Estimation of Intra-class Correlation Parameter for Correlated Binary Data In Common Correlated Models

Zhang Hao

(B.Sc. Peking University)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgements

For the completion of this thesis, I would like very much to express my heartfelt gratitude to my supervisor Associate Professor Yougan Wang for all his invaluable advice and guidance, endless patience, kindness and encouragement during the past two years. I have learned many things from him regarding academic research and character building.

I also wish to express my sincere gratitude and appreciation to my other lecturers, namely Professors Zhidong Bai, Zehua Chen and Loh Wei Liem, etc., for imparting knowledge and techniques to me and their precious advice and help in my study.

It is a great pleasure to record my thanks to my dear friends: to Ms. Zhu Min, Mr. Zhao Yudong, Mr. Ng Wee Teck, and Mr. Li Jianwei for their advice and help in my study; to Mr. and Mrs. Rong, Mr. and Mrs. Guan, Mr. and Mrs. Xiao, Ms. Zou Huixiao, Ms. Peng Qiao and Ms. Qin Xuan for their kind help and warm encouragement in my life during the past two years.

Finally, I would like to attribute the completion of this thesis to other members and staff of the department for their help in various ways and providing such a pleasant working environment, especially to Jerrica Chua for administrative matters and Mrs. Yvonne Chow for advice in computing.

Zhang Hao

July, 2005

Contents

1 Introduction **2**

1.1 Common Correlated Model 2

1.2 Two Specifications of the Common Correlated Model 5

1.2.1 Beta-Binomial Model 5

1.2.2 Generalized Binomial Model 6

1.3 Application Areas 7

1.3.1 Teratology Study 7

1.3.2 Other Uses 9

1.4 The Review of the Past Work 10

1.5 The Organizations of the Thesis 11

2 Estimating Equations **12**

2.1 Estimation for the mean parameter π 12

2.2 Estimation for the ICC ρ 14

2.2.1 Likelihood based Estimators 14

2.2.2 Non-Likelihood Based Estimators 16

2.3 The Past Comparisons of the Estimators 26

2.4 The Estimators We Compare 27

2.5 The Properties of the Estimators 28

2.5.1 The Asymptotic Variances of the Estimators 28

2.5.2 The Relationship of the Asymptotic Variances 39

3 Simulation Study 41

3.1 Setup 41

3.2 Results 45

3.2.1 The Overall Performance 45

3.2.2 The Effect of the Various Factors 48

3.2.3 Comparison Between Different Estimators 49

3.3 Conclusion 52

4 Real Examples 62

4.1 The Teratological Data Used in Paul 1982 62

4.2 The COPD Data Used in Liang 1992 62

4.3 Results 63

5 Future Work 66

Summary

In common correlation models, the intra-class correlation parameter (ICC) provides a quantitative measure of the similarity between individuals within the same cluster. The estimation for ICC parameter is of increasing interest and important use in biological and toxicological studies, such as the disease aggression study and the Teratology study.

The thesis mainly compares the following four estimators for the ICC parameter ρ: the Kappa-type estimator (ρ_{FC}), the Analysis Of Variance estimator (ρ_A), the Gaussian likelihood estimator (ρ_G) and a new estimator (ρ_{UJ}) that is based on the Cholesky Decomposition. The new estimator is a specification of the UJ method proposed by Wang and Carey (2004) and has not been considered before.

Analytic expressions of the asymptotic variances of the four estimators are obtained and extensive simulation studies are carried out. The bias, standard deviation, the mean square error and the relative efficiency for the estimators are compared. The results show that the new estimator performs well when the mean and correlation are small.

Two real examples are used to investigate and compare the performance of these estimators in practice.

keyword: *binary clustered data analysis, common correlation model, intra-class correlation parameter/coefficient, Cholesky Decomposition, Teratology study*

List of Tables

1.1 A Typical Data in Teratological Study (Weil, 1970) 8

3.1 Distributions of the Cluster Size 43

3.2 The effect of various factors on the bias of the estimator ρ_{UJ} in 1000 simulations from a beta binomial distribution. 53

3.3 The effect of various factors on the mean square error of ρ_{UJ} in 1000 simulations from a beta binomial distribution. 54

3.4 The MSE of ρ_{FC} and ρ_{UJ} when the cluster size distribution is Kupper . 55

3.5 The MSE of ρ_{FC} and ρ_{UJ} when the cluster size distribution is Brass . . 55

3.6 The "turning point" of ρ when $\pi = 0.05$ 55

4.1 Shell Toxicology Laboratory, Teratology Data 63

4.2 COPD familial disease aggregation data 63

4.3 Estimating Results for the Real Data Sets 64

4.4 The Estimated value of the Asymptotic Variance of $\hat{\rho}$ (By plugging the estimates of (π, ρ)) into formulas: (2.29), (2.28), (2.26) and (2.21) . . . 65

4.5 The Estimated value of the Asymptotic Variance of $\hat{\rho}$ (by using the Robust Method) 65

List of Figures

3.1 The two distributions of the cluster size n_i 44

3.2 The overall performances of the four estimators when $k = 10$. 46

3.3 The overall performances of the four estimators when $k = 25$ 47

3.4 The overall performances of the four estimators when $k = 50$ 48

3.5 The Legend for Figure (3.8), (3.7), (3.6), (3.9) and (3.10) 56

3.6 The Relative Efficiencies when $k = 25$ and $\pi = 0.5$ 57

3.7 The Relative Efficiencies when $k = 25$ and $\pi = 0.2$ 58

3.8 The Relative Efficiencies when $k = 25$ and $\pi = 0.05$ 59

3.9 The Relative Efficiencies when $k = 10$ and $\pi = 0.05$ 60

3.10 The Relative Efficiencies when $k = 50$ and $\pi = 0.05$ 61

Chapter 1

Introduction

1.1 Common Correlated Model

Data in the form of clustered binary response arise in the toxicological and biological studies in the recent decades. Such kind of data are in the form like this: there are several identical individuals in one cluster and the response for each individual is dichotomous. For ease of the presentation, we name the binary responses here as "alive" or "dead", and the metric (0,1) is imposed with 0 for "alive" and 1 for "dead".

Suppose there are n_i individuals in the i^{th} cluster and there are k clusters in total. The binary response for the j^{th} individual in the i^{th} cluster is denoted as $y_{ij} = 1/0 \quad (i = 1, 2, ..., k; j = 1, 2, ..., n_i)$. So $S_i = \sum_{j=1}^{n_i} y_{ij}$ is the total number of the individuals observed to respond 1 in the i^{th} cluster. It is postulated that the "death" rate of all the individuals in the i^{th} cluster are the same, which is $P(y_{ij} = 1) = \pi$. The correlation between any two individuals in the same cluster are assumed to be the same. We denote this Intra-Class Correlation parameter as $\rho = \text{Corr}(y_{il}, y_{ik})$ for any $l \neq k$. For individuals from different clusters, they are assumed to be independent, which means y_{ij} is independent of y_{mn} for any $i \neq m$.

The variance of S_i often exhibit greater value than the predicted value if a simple binomial model is used. This phenomenon is called the over-dispersion, which is due to the tendency that the individuals in the same cluster would respond more likely than individuals from different clusters.

According to the above assumptions, we can see that:

$$\text{E}y_{ij} = \pi \quad \textit{and} \quad \text{Var}y_{ij} = \pi(1-\pi) \quad i = 1, 2, \ldots k \quad j = 1, 2, \ldots n_i$$

And for the sum variable $S_i = \sum_{j=1}^{n_i} y_{ij}$, which is the sufficient statistics for π:

$$\text{E}S_i = n_i\pi \quad \textit{and} \quad \text{Var}S_i = n_i\pi(1-\pi)(1 + (n_i - 1)\rho)$$

The second moment of S_i is determined by ρ but the third, forth and the higher order moment of S_i may depend on the other parameters. Only when we know the likelihood of S_i (such as the Beta-binomial model or the generalized binomial model), we can get the closed forms of these higher order moment of S_i.

Define a series of parameters:

$$\phi_s = \frac{\text{E}\prod_{j=1}^{j=s}(y_{ij} - \pi)}{\text{E}(y_{i1} - \pi)^s} \quad s = 2, 3, \ldots$$

For the common correlated model, we can show that $\phi_2 = \rho$ and the s^{th} moment $m_{si} = \text{E}(S_i - n_i\pi)^s$ of S_i only depends on $\{\pi, \phi_2, \ldots, \phi_s\}$

When π is fixed, ρ can not take all the values between $(-1, 1)$. Prentice(1986) has given the general constraints for the binary response model:

$$\rho \geq \frac{-1}{n_{max}-1} + \frac{\omega(1-\omega)}{n_{max}(n_{max}-1)\pi(1-\pi)}$$

where $n_{max} = max\{n_1, n_2, \dots, n_k\}$, $\omega = n_{max}\pi - int(n_{max}\pi)$ and $int(.)$ means the integer part of any real number. For the different specifications of the model, the constraints might be different.

The model described above was first formally suggested as the **Common Correlated Model** by Landis and Koch (1977a). It includes various specifications, such as Beta-Binomial and Extended Beta-Binomial model (BB) of Crowder (1986), Correlated Beta-Binomial model (CB) of Kupper and Haseman (1978) and the Generalized Binomial model (GB) of Madsen (1993).

Kupper and Haseman (1978) has given an alternative specification of the common correlated model when ρ is positive. It is assumed that the probability of alive (success) varies from group to group (but keep the same between individuals in the same group) according to a distribution with mean π and variance $\rho\pi(1-\pi)$. All the individuals (both within the same group and different groups) are independent conditional on this probability. If this probability is distributed according to Beta distribution, it will lead to the well-known Beta-Binomial model.

1.2 Two Specifications of the Common Correlated Model

1.2.1 Beta-Binomial Model

Of the specifications of the common correlated model, Beta-Binomial model is the most popular. Paul (1982) and Pack (1986) has shown the superiority of the beta-binomial model for the analysis of proportions. However, Feng and Grizzle (1992) found that the BB model is too restrictive to be relied on for inference when n_i are variable.

The beta-binomial distribution is derived as a mixture distribution in which the probability of alive varies from group to group according to a beta distribution with parameters α and β. S_i is binomially distributed conditional on this probability. In terms of the parameterizations of α and β, the marginal probability of alive for any individual is: $\pi = \alpha/(\alpha + \beta)$ and the intra-class correlation parameter is: $\rho = 1/(1 + \alpha + \beta)$. Denote $\theta = 1/(\alpha + \beta)$, we can get the probability function for the Beta-Binomial Distribution:

$$\begin{aligned} P(S_i = y) &= \binom{n_i}{y} \frac{B(\alpha + y, n_i + \beta - y)}{B(\alpha, \beta)} \\ &= \binom{n_i}{y} \frac{\prod_{j=0}^{y-1}(\pi + j\theta) \prod_{j=0}^{n_i-y-1}(1 - \pi + j\theta)}{\prod_{j=0}^{y-1}(1 + j\theta)} \end{aligned} \tag{1.1}$$

If the intra-class correlation $\rho > 0$, it is called over-dispersion, otherwise it is called under-dispersion. Over-dispersion is much more common than under-dispersion in practice since the litter effect suggests that any two individuals are tended to respond more likely and therefore they are positively correlated. But this does not mean that ρ must be positive. For BB model, it is required that $\rho > 0$. However, Crowder (1986) showed that to ensure (1.1) to be a probability function, ρ only needs to satisfy

$$\rho > -min\{\frac{\pi}{n_{max} - 1 - \pi}, \frac{1-\pi}{n_{max} - 1 - (1-\pi)}\}$$

In this case, ρ can take negative values, which makes the BB model also suitable for under-dispersion data. This is called extended beta-binomial model.

1.2.2 Generalized Binomial Model

The generalized binomial model is proposed by Madsen (1993). It can be treated as the mixture of two binomial distributions:

$$Y = \rho X_1 + (1-\rho) X_2$$

Where

$$P(X_1 = x) = \begin{cases} 1-\pi & x = 0 \\ \pi & x = n \end{cases} \qquad and \quad X_2 \sim Binomial(n, \pi)$$

So the probability can be written down as:

$$P(Y = y) \quad = \quad \begin{cases} \rho(1-\pi) + (1-\rho)(1-\pi)^n, & y = 0 \\ (1-\rho)\binom{n}{y}\pi^y(1-\pi)^{n-y}, & 1 \le y \le n-1 \\ \rho\pi + (1-\rho)\pi^n, & y = n \end{cases} \tag{1.2}$$

To ensure (1.2) to be a probability mass function, the constraint for ρ is:

$$max\{-\frac{(1-\pi)^n}{(1-\pi)-(1-\pi)^n}, -\frac{\pi^n}{\pi - \pi^n}\} \leqslant \rho \leqslant 1$$

An advantage of the generalized binomial model is that ρ contains information for the higher(≥ 3) order moment. As we know, the correlation for any pair

$$\mathrm{Corr}(y_{ij}, y_{ik}) = \frac{\mathrm{E}(y_{ij}-\pi)(y_{ik}-\pi)}{\mathrm{E}(y_{ij}-\pi)^2} = \rho = \phi_2$$

For the GB model, it can be shown that:

$$\begin{aligned}
\frac{\mathrm{E}(y_{ij}-\pi)(y_{ik}-\pi)(y_{il}-\pi)}{\mathrm{E}(y_{ij}-\pi)^3} &= \phi_3 = \rho \\
\frac{\mathrm{E}(y_{ij}-\pi)(y_{ik}-\pi)(y_{il}-\pi)(y_{im}-\pi)}{\mathrm{E}(y_{ij}-\pi)^4} &= \phi_4 = \rho
\end{aligned}$$

That means ρ also determines the third and forth moment of S_i.

1.3 Application Areas

1.3.1 Teratology Study

Of the various applied areas of the common correlated model, we mainly focus on the Teratology studies. In a typical Teratology study, female rats are exposed to different dose of drugs when they are pregnant. Each fetus is examined and a dichotomous response variable indicating the presence or absence of a particular response (e.g., malformation) is recorded. For ease of the presentation, we often denote the dichotomous response as alive or dead. Apply the common correlation model and the notations above to the teratology study, it can be described as: k female rats were exposed to certain dose of drug during their pregnancy. For the i^{th} rat, she gave birth to n_i fetuses. Of the n_i fetuses, y_{ij} denotes the survival status for the j^{th} fetus. $y_{ij} = 1$ means the fetus is observed dead or it is alive. Then $S_i = \sum_{j=1}^{n_i} y_{ij}$ is the total number of fetuses that are observed to be dead out of all the n_i fetuses given birth by the i^{th} female rat.

Here is an example of the data that appeared in a typical Teratology study. The data below are from a teratological experiment comprised of two treatments ("two dose") by Weil (1970). Sixteen pregnant female rats were fed a control diet during pregnancy and lactation, whereas an additional 16 were treated with a chemical agent. Each proportion represents the number of pups that survived the 21-day lactation period among those who were alive at 4 days.

Table 1.1: A Typical Data in Teratological Study (Weil, 1970)

i	*Control* (n_i/S_i)	*Treated* (n_i/S_i)
1	13/13	12/12
2	12/12	11/11
3	9/9	10/10
4	9/9	9/9
5	8/8	11/10
6	8/8	10/9
7	13/12	10/9
8	12/11	9/8
9	10/9	9/8
10	10/9	5/4
11	9/8	9/7
12	13/11	7/4
13	5/4	10/5
14	7/5	6/3
15	10/7	10/3
16	10/7	7/0

It can be shown that only 25% of the total sample variation from the treated group can be accounted for by binomial variation (Liang and Hanfelt, 1994). This is a typical over-dispersion clustered binary response data and the ICC parameter ought to be positive.

1.3.2 Other Uses

Besides the Teratological studies, the estimation for the intra-class correlation coefficient are also widely used in the other fields of toxicological and biological studies. For example, Donovan, Ridout and James (1994) used the ICC to quantify the extent of variation in rooting ability among somaclones of the apple cultivar Greensleeves; Gibson and Austin (1996) used an estimator of ICC to characterize the spatial pattern of disease incidence in an orchard; Barto (1966), Fleiss and Cuzick (1979) and Kraemer et al.(2002) used ICC as an index measuring the level of interobserver agreement; Gang et al. (1996) used ICC to measure the efficiency of hospital staff in the health delivery research; Cornfield (1978) used ICC for estimating the required size of a cluster randomization trial.

In some clustered binary situation, the ICC parameter can be interpreted as the "heritability of a dichotomous trait" (Crowder 192, Elston, 1977). It is also frequently used to quantify the familial aggregation of disease in the genetic epidemiological studies (Cohen, 1980; Liang, Quaqish and Zeger, 1992).

1.4 The Review of the Past Work

Donner (1986) has given a summarized review for the estimators of ICC in the case that the responses are continuous. He also remarked that the application of continuous theory for the binary response has severe limitations. In addition, the moment method to estimate the correlation, which is used in the GEE approach proposed by Liang and Zegger (1986) is also not appropriate for the estimation of ICC when the response is binary.

A commonly used method to estimate ICC is the Maximum likelihood method based on the Beta-Binomial model (Williams 1975) or the extended beta binomial model (Prentice 1986). However the estimator based on the parametric model may yield inefficient or biased results when the true model was wrongly specified.

Some robust estimators which are independent of the distributions of S_i have been introduced, such as the moment estimator (Klienman, 1973), analysis of variance estimator (Eslton, 1977), quasi-likelihood estimator (Breslow, 1990; Moore and Tsiatis, 1991), extended quasi-likelihood estimator (Nelder and Pregibon, 1987), pseudo-likelihood estimator (Davidian and Carroll, 1987) and the estimators based on the quadratic estimating equations (Crowder 1987; Godambe and Thompson 1989).

Ridout et al. (1999) had given an excellent review of the earlier works and conducted a simulation study to compare the bias, standard deviation, mean square error and the relative efficiencies of 20 estimators. The reviewing work is based on the data simulated from beta binomial and mixture binomial distributions and the simulation results showed that seven estimators performed well as far as these properties were concerned. Paul (2003) introduced 6 new estimators based on the quadratic estimating equations and compare these estimators along with the 20 estimator used by Ridout et al. (1999). Paul's work shows that an estimator based on the quadratic estimating equations also perform well for the joint estimation of (π, ρ).

1.5 The Organizations of the Thesis

Chapter 1(this chapter) gives an introduction to the clustered binary data, common correlated model and reviews the past works on the estimation of the ICC ρ. Chapter 2 will introduce the commonly used estimators and the new estimators that we are going to investigate. Then we will obtain the asymptotic variances of the four estimators that we are going to compare: κ-type (FC) estimator, ANOVA estimator, Gaussian likelihood estimator and the new estimator based on Cholesky decomoposition. Chapter 3 will carry the simulation studies to compare the performances of these four estimators. We will compare the bias, standard deviation, mean square error and the relative efficiency of these four estimators. To investigate the performance of the estimators in practice, chapter 4 will apply these four estimators on two real example data sets. Chapter 5 will give general conclusions and describe the future work.

Chapter 2

Estimating Equations

2.1 Estimation for the mean parameter π

Since S_i is the sufficient statistics for π, modelling on the vector response y_{ij} does not give more information for π than modelling on $S_i = \sum_{j=1}^{n_i} y_{ij}$. On the other hand, the estimating equation should not dependent on the order of the fetuses in the developmental studies. Denote the residual $g_i = S_i - n_i\pi$ and the variance $V_i = \mathrm{Var}(S_i - n_i\pi) = \sigma_i^2 = n_i\pi(1-\pi)[1+(n_i-1)\rho]$. Use the Quasi-likelihood approach, we can get the estimating equation for π:

$$\begin{aligned}
U(\pi;\rho) &= \sum_{i=1}^{k} D_i V_i^{-1} g_i \\
&= \sum_{i=1}^{k} -\frac{\partial(S_i - n_i\pi)}{\partial\pi}\sigma_i^{-2}(S_i - n_i\pi) \\
&= \sum_{i=1}^{k} \frac{S_i - n_i\pi}{\pi(1-\pi)[1+(n_i-1)\rho]} \qquad (2.1)
\end{aligned}$$

Simplify (2.1), we get the Quasi-likelihood estimating equation for π:

$$\begin{aligned} U(\pi;\rho) &= \sum_{i=1}^{k} \frac{S_i - n_i\pi}{1+(n_i-1)\rho} \\ &= \sum_{i=1}^{k} \frac{S_i - n_i\pi}{\nu_i}, \end{aligned} \tag{2.2}$$

Where $\nu_i = 1 + (n_i - 1)\rho$

From another point of view,we may also use the GEE approach, which is modelled on the vector response $y_i = \{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$.

$$U(\pi;\rho) = \sum_{i=1}^{k} \mathbf{1_{n_i}}^{\mathrm{T}} V_i^{-1} \left\{ \begin{pmatrix} y_{i1} \\ y_{i2} \\ \ldots \\ y_{in_i} \end{pmatrix} - \pi \begin{pmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{pmatrix} \right\}$$

where $\mathbf{1_{ni}}$ is the vector consisting of ones, $V_i = \mathrm{Cov}(Y_i) = \pi(1-\pi)[(1-\rho)I + \rho\mathbf{1}\mathbf{1}^{\mathrm{T}}]$. Thus

$$V_i^{-1} = \frac{1}{\pi(1-\pi)(1-\rho)}\{I - \frac{\rho}{1+(n_i-1)\rho}\mathbf{1}\mathbf{1}^{\mathrm{T}})\}.$$

Then the GEE estimating equation for π can be written as:

$$U(\pi;\rho) = \sum_{i=1}^{k} \frac{(S_i - n_i\pi)}{\pi(1-\pi)[1+(n_i-1)\rho]} \tag{2.3}$$

Note that (2.3) also does not depend on the order of y_{ij} even though it is modelled on the vector response. It has the same form with the Quasi-likelihood estimating equation (2.1).

Consider a general set of estimators for π:

$$\hat{\pi} = \frac{\sum_i \omega_i S_i}{\sum_i \omega_i n_i} \tag{2.4}$$

When $w_i = [1 + (n_i - 1)\rho]^{-1} = \nu_i^{-1}$, we can get (2.2). The weight factor ω_i can also take other values. For example, when $\omega_i = 1$, the estimator for π is $\hat{\pi} = \sum_i S_i / \sum_i n_i$ and when $\omega = 1/n_i$, the estimator for π is $(\sum_i S_i/n_i)/k$

2.2 Estimation for the ICC ρ

2.2.1 Likelihood based Estimators

The maximum likelihood estimators are based on the parametric models. However, when the parametric model does not fit the data well, these estimators may be highly biased or inefficient.

- MLE Estimator Based on Beta Binomial Model

  As mentioned in (1.2.1), the likelihood of the beta binomial distribution is:

$$
\begin{aligned}
P(S_i = y) &= \binom{n_i}{y} \frac{B(\alpha + y, n_i + \beta - y)}{B(\alpha, \beta)} \\
&= \binom{n_i}{y} \frac{\prod_{j=0}^{y-1}(\pi + j\theta) \prod_{j=0}^{n_i-y-1}(1 - \pi + j\theta)}{\prod_{j=0}^{y-1}(1 + j\theta)}
\end{aligned}
$$

  Denote the log-likelihood as $l(\pi, \rho)$, so the jointly estimating equations for (π, ρ) is:

$$
\frac{\partial l}{\partial \pi} = \sum_{i}^{k} \left\{ \sum_{r=0}^{S_i-1} \frac{1-\rho}{(1-\rho)\pi + r\rho} - \sum_{r=0}^{n_i-S_i-1} \frac{1-\rho}{(1-\rho)(1-\pi) + r\rho} \right\} = 0
$$

  and

$$
\frac{\partial l}{\partial \rho} = \sum_{i=1}^{k} \left\{ \sum_{r=1}^{S_i-1} \frac{\rho - \pi}{(1-\rho)\pi + r\rho} + \sum_{r=0}^{n_i-S_i-1} \frac{r - (1-\pi)}{(1-\rho)(1-\pi) + r\rho} - \sum_{r=0}^{n_i-1} \frac{r-1}{(1-\rho) + r\rho} \right\} = 0
$$

Denote the solution for the above estimating equations as the maximum likelihood estimator ρ_{ML}

- Gaussian Likelihood Estimator

The Gaussian likelihood estimator was introduced by Whittle (1961) when dealing with the continuous response and Crowder(1985) introduced it to the analysis of binary data. As shown in Chapter 1, we know that the Gaussian likelihood model only needs to assume the first two moments and are very easy to calculate of all the moment based methods. Paul (2003) also showed that the Gaussian estimator for the binary data performance well, compared with the other known estimators for ICC.

Assume the vector response $y_i = \{y_{i1},\ y_{i2},\ \ldots,\ y_{in_i}\}$ is distributed according to the multivariate Gaussian distribution, with the mean and variance:

$$\mathrm{E}y_i = \tilde{\mu} = \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix} \text{ and } \mathrm{Var}(y_i) = \begin{pmatrix} 1 & \rho & \rho & \ldots & \rho \\ \rho & 1 & \rho & \ldots & \rho \\ & & \ldots & & \\ \rho & \rho & \ldots & 1 & \rho \\ \rho & \rho & \ldots & \rho & 1 \end{pmatrix} = A_i^{1/2} R_i A_i^{1/2}$$

Here $A_i = diag\{\pi(1-\pi), \pi(1-\pi), \ldots, \pi(1-\pi)\}$ is the diagonal variance matrix. Denote the residual

$$\varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \ldots \\ \varepsilon_{in_i} \end{pmatrix} = \begin{pmatrix} y_{i1} - \pi \\ y_{i2} - \pi \\ \ldots \\ y_{in_i} - \pi \end{pmatrix}$$

the standardized residual

$$\epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \cdots \\ \epsilon_{in_i} \end{pmatrix} = \begin{pmatrix} \frac{y_{i1}-\pi}{\sqrt{\pi(1-\pi)}} \\ \frac{y_{i2}-\pi}{\sqrt{\pi(1-\pi)}} \\ \cdots \\ \frac{y_{in_i}-\pi}{\sqrt{\pi(1-\pi)}} \end{pmatrix} = A_i^{-1/2}\varepsilon_i$$

and $l(\pi, \rho)$ to be the log-likelihood of Gaussian distribution.

So $-2 * l(\pi, \rho) = log|A_i^{-1/2} R_i^{-1} A_i^{-1/2}| + \epsilon_i^T R_i \epsilon_i$. Let $\dfrac{\partial(-2 * l(\pi, \rho))}{\partial \rho} = 0$, we have:

$$\begin{aligned} U_G^* &= \sum_i \left(\epsilon_i^T \frac{\partial R_i^{-1}}{\partial \rho} \epsilon_i - tr(\epsilon_i^T R_i^{-1} \epsilon_i) \right) \\ &= \sum_i \left(\epsilon_i^T \frac{\partial R_i^{-1}}{\partial \rho} \epsilon_i - tr(\frac{\partial R_i^{-1}}{\partial \rho} R_i) \right) \\ &= \sum_i \frac{\{\rho(n_i-1)[2+(n_i-2)\rho] \sum_l \epsilon_{il}^2 - (1+(n_i-1)\rho^2) \sum_{l \neq k} \epsilon_{il}\epsilon_{ik}\}}{(1-\rho)^2[1+(n_i-1)\rho]^2} \\ &= \sum_i \frac{\{(1-2\pi)[1+(n_i-1)\rho]^2(S_i - n_i\pi) - [1+(n_i-1)\rho^2][(S_i-n_i\pi)^2 - m_{2i}]\}}{(1-\rho)^2[1+(n_i-1)\rho]^2\pi(1-\pi)} \end{aligned}$$

To simplify U_G^*, we can get the Gaussian estimating equation as:

$$U_G = \sum_i \left\{ (1-2\pi)(S_i - n_i\pi) - \frac{1+(n_i-1)\rho^2}{[1+(n_i-1)\rho]^2} \left[(S_i - n_i\pi)^2 - m_{2i} \right] \right\} \quad (2.5)$$

Denote the solution for (2.5) as the Gaussian likelihood estimator ρ_G.

2.2.2 Non-Likelihood Based Estimators

The non likelihood based estimators are supposed to be more robust than the maximum likelihood estimators since they are independent of the distributions of S_i. We will introduce the new estimator ρ_{UJ} which based on the Cholesky decomposition, as well as some other commonly used estimators.

- New Estimator Based on Cholesky Decomposition

The new estimator is a specification of the U-J method proposed by Wang and Carey (2004), which is based on the Cholesky Decomposition:

$$U_J = \sum_i \varepsilon_i^T \frac{\partial B_i^T}{\partial \rho} J_i B_i \varepsilon_i \qquad where \qquad \varepsilon_{il} = y_{il} - \pi \quad and \quad R_i^{-1} = B_i^T J_i B_i$$

Here B_i is a lower triangular matrix with the leading value of 1 and J_i is a diagonal matrix.

Since R_i is the compound symmetry correlation matrix, we have:

$$R_i = (1-\rho)I + \rho 1_{n_i} 1'_{n_i} \quad and \quad R_i^{-1} = \frac{1}{1-\rho} I - \frac{\rho}{(1-\rho)[1+(n_i-1)\rho]} 1_{n_i} 1'_{n_i}$$

So the lower triangular matrix B_i and diagonal matrix J_i can be written as:

$$Bi = \begin{pmatrix} 1 & & & & \\ -\rho & 1 & & & \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 & & \\ & & \vdots & \ddots & \\ -\frac{\rho}{1+(n_i-2)\rho} & & \cdots & -\frac{\rho}{1+(n_i-2)\rho} & 1 \end{pmatrix}$$

and

$$J_i = diag\left\{ \frac{1+(j-2)\rho}{(1-\rho)[1+(j-1)\rho]} \right\}$$

So,

$$\frac{\partial B_i^T}{\partial \rho} = \begin{pmatrix} 0 & -1 & -\frac{1}{(1+\rho)^2} & -\frac{1}{(1+2\rho)^2} & \cdots & -\frac{1}{(1+(n_i-2)\rho)^2} \\ & 0 & -\frac{1}{(1+\rho)^2} & -\frac{1}{(1+2\rho)^2} & \cdots & -\frac{1}{(1+(n_i-2)\rho)^2} \\ & & 0 & \cdots & & \vdots \\ & & & & & 0 \end{pmatrix}$$

and

$$\varepsilon_i^T \frac{\partial B_i^T}{\partial \rho} = \begin{pmatrix} \varepsilon_{i1} & \cdots & \varepsilon_{in_i} \end{pmatrix} \begin{pmatrix} 0 & -1 & -\frac{1}{(1+\rho)^2} & -\frac{1}{(1+2\rho)^2} & \cdots & -\frac{1}{(1+(n_i-2)\rho)^2} \\ & 0 & -\frac{1}{(1+\rho)^2} & -\frac{1}{(1+2\rho)^2} & \cdots & -\frac{1}{(1+(n_i-2)\rho)^2} \\ & & 0 & \cdots & & \vdots \\ & & & & & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0, & \cdots & -\frac{1}{[1+(j-2)\rho]^2} \sum_{l=1}^{j-1} \varepsilon_{il} & \cdots & -\frac{1}{(1+(n_i-2)\rho)^2} \sum_{l=1}^{n_i-1} \varepsilon_{il} \end{pmatrix}$$

$$B_i \varepsilon_i = \begin{pmatrix} 1 & & & & \\ -\rho & 1 & & & \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 & & \\ & & \vdots & \ddots & \\ -\frac{\rho}{1+(n_i-2)\rho} & & & -\frac{\rho}{1+(n_i-2)\rho} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ -\frac{\rho}{1+(j-2)\rho} \sum_{l=1}^{j-1} \varepsilon_{il} + \varepsilon_{ij} \\ \vdots \\ -\frac{\rho}{1+(n_i-2)\rho} \sum_{l=1}^{n_i-1} \varepsilon_{il} + \varepsilon_{in_i} \end{pmatrix}$$

Thus:

$$\begin{aligned} \sum_i \varepsilon_i^T \frac{\partial B_i^T}{\partial \rho} J_i B_i \varepsilon_i &= \sum_i \sum_{j=2}^{n_i-2} \frac{1+(j-2)\rho}{(1-\rho)[1+(j-1)\rho]} \left[-\frac{1}{[1+(j-2)\rho]^2} \sum_{l=1}^{j-1} \varepsilon_{il} \right] \\ &\times \left[-\frac{\rho}{1+(j-2)\rho} \sum_{l=1}^{j-1} \varepsilon_{il} + \varepsilon_{ij} \right] \end{aligned}$$

$$U_{Ji} = \frac{1}{1-\rho} \sum_i \sum_{j=2}^{n_i} \left[\frac{\varepsilon_{ij} \sum_{l=1}^{j-1} \varepsilon_{il}}{[1+(j-1)\rho][1+(j-2)\rho]} - \rho \frac{(\sum_{l=1}^{j-1} \varepsilon_{il})^2}{[1+(j-1)\rho][1+(j-2)\rho]^2} \right]$$

Let's consider all the permutations of ε_{ij}. We use $\varepsilon_{i[l]}$ represent one permutation. Since there are $n_i!$ permutations for the i^{th} cluster, we shall use $1/n_i!$ as the weight for the i^{th} cluster.

$$
\begin{aligned}
U_J &= \frac{1-\rho}{n_i!} U_{Ji} \\
&= \sum_i \frac{\sum_{P[1,2\ldots n_i]} \sum_{j=2}^{n_i} \left[\frac{\varepsilon_{i[j]} \sum_{l=1}^{j-1} \varepsilon_{i[l]}}{[1+(j-1)\rho][1+(j-2)\rho]} - \rho \frac{(\sum_{l=1}^{j-1} \varepsilon_{i[l]})^2}{[1+(j-1)\rho][1+(j-2)\rho]^2} \right]}{n_i!} \\
&= \sum_i \sum_{j=2}^{n_i} \frac{j-1}{[1+(j-1)\rho][1+(j-2)\rho]^2} \left[\rho \frac{\sum_l \varepsilon_{il}^2}{n_i} - \frac{\sum_{l\neq k} \varepsilon_{il}\varepsilon_{ik}}{n_i(n_i-1)} \right] \\
&= \sum_i \frac{C_i}{n_i(n_i-1)} \left[\rho(n_i-1) \sum_l \varepsilon_{il}^2 - \sum_{l\neq k} \varepsilon_{il}\varepsilon_{ik} \right]
\end{aligned}
$$

where $C_i = \sum_{j=2}^{n_i} \dfrac{j-1}{[1+(j-1)\rho][1+(j-2)\rho]^2}$

Since

$$
\begin{aligned}
\sum_l \varepsilon_{il}^2 &= \sum_l (y_{il}-\pi)^2 \\
&= \sum_l (y_{il}^2 - 2\pi y_{il} + \pi^2) = \sum_l [(y_{il}(1-2\pi)) + \pi^2] \\
&= (S_i - n_i\pi)(1-2\pi) + n_i\pi(1-\pi)
\end{aligned}
$$

and

$$
\sum_{l\neq k} \varepsilon_{il}\varepsilon_{ik} = (S_i - n_i\pi)^2 - [(S_i - n_i\pi)(1-2\pi) + n_i\pi(1-\pi)]
$$

Hence we can get a more easy to calculate form of the U-J estimating equation:

$$
U_J = \sum_i \frac{C_i}{n_i(n_i-1)} \left[(1-2\pi)[1+(n_i-1)\rho](S_i - n_i\pi) - (S_i - n_i\pi)^2 + m_{2i} \right] \tag{2.6}
$$

where

$$C_i = \sum_{j=2}^{n_i} \frac{j-1}{[1+(j-1)\rho][1+(j-2)\rho]^2}$$

Denote the solution for (2.6) as the new estimator ρ_{UJ}

- The Analysis of Variance Estimator

The analysis of variance estimator is by defination:

$$\hat{\rho}_A = \frac{MSB - MSW}{MSB + (n_A - 1)MSW} \tag{2.7}$$

where MSB and MSW are the between and within group mean squares from a one-way analysis of variance of the response y_i. And

$$n_A = \frac{1}{k-1}\left(N - \frac{\sum n_i^2}{N}\right) \quad \text{and} \quad N = \sum_i n_i$$

The analysis of variance estimator was first used to deal with the continuous response. Elston (1977) firstly suggest to use it for the binary responses. For the binary data, the mean squares MSB and MSW are defined as:

$$MSB = \frac{1}{k-1}\left\{\sum \frac{Y_i^2}{n_i} - \frac{(\sum Y_i)^2}{N}\right\}, \quad MSW = \frac{1}{N-k}\left\{\sum Y_i - \sum \frac{Y_i^2}{n_i}\right\}$$

- Direct Probability Interpretation Estimators

Assume the probability that two individuals have the same response to be α if they are from the same cluster or β if they are from the different clusters. The assumptions of the common correlation model shows that:

$$\alpha = 1 - 2\pi(1-\pi)(1-\rho) \quad \text{and} \quad \beta = 1 - 2\pi(1-\pi)$$

and hence that

$$\rho = \frac{\alpha - \beta}{1 - \beta} \tag{2.8}$$

The estimators based on the direct probability interpretation is by evaluating α and β in (2.8) with their estimators.

If we estimate α as a weighted average of $\alpha_i = 1 - \frac{2S_i(n_i - S_i)}{n_i(n_i-1)}$, with weights proportional to $n_i - 1$ and estimate β as $1 - 2\hat{\pi}(1-\hat{\pi})$, where $\hat{\pi} = \frac{\sum_i S_i}{\sum_i n_i}$, we can obtain the κ-type estimators proposed by Fleiss and Cuzick (1979):

$$\rho_{FC} = 1 - \frac{1}{(N-k)\hat{\pi}(1-\hat{\pi})} \sum_i \frac{S_i(n_i - S_i)}{n_i}, \qquad \hat{\pi} = \frac{\sum_i S_i}{\sum_i n_i} \tag{2.9}$$

Similarly, we can get other estimators with the different estimators of α and β. Mak (1988) has proposed the Mak's estimator:

$$\rho_{MAK} = 1 - \frac{(k-1)\sum_i \frac{S_i(n_i - S_i)}{n_i(n_i-1)}}{\sum_i \frac{S_i^2}{n_i^2} + (\sum_i \frac{S_i}{n_i})(k - 1 - \sum_i \frac{S_i}{n_i})} \tag{2.10}$$

Mak (1988) shown that: for the beta binomial data, these two estimators (ρ_{FC} and ρ_{Mak}) are better than the other estimators that are based on probability interpretation.

- Direct Calculation of Correlation Estimator

Donner (1986) suggested to estimate ρ by calculating the Pearson correlation coefficient over all possible pairs within one group. Karlin et al. (1981) proposed the general form of such kind of estimators. Ridout et al. (1999) extended this method to the binary data and proposed the Pearson correlation estimator as:

$$\rho_{PW} = \frac{\sum_i \omega_i S_i(S_i - 1) - \hat{\pi}^2}{\hat{\pi}(1 - \hat{\pi})} \tag{2.11}$$

where

$$\hat{\pi} = \sum_i \omega_i (n_i - 1) S_i \quad \text{and} \quad \sum_i n_i(n_i - 1)\omega_i = 1$$

Ridout et al. (1999) used three weights

$$\omega_i = \frac{1}{\sum_i n_i(n_i - 1)} \quad , \quad \omega_i = \frac{1}{kn_i(n_i - 1)} \quad \text{and} \quad \omega_i = \frac{1}{N(n_i - 1)}$$

Denote the estimator that use the constant weight $\omega_i = 1/\sum_i n_i(n_i - 1)$ as the Pearson estimator $\rho_{Pearson}$.

$$\rho_{Pearson} = \frac{1}{\hat{\pi}(1 - \hat{\pi})}\left[\frac{\sum_i S_i(S_i - 1)}{n_i(n_i - 1) - \hat{\pi}^2}\right] \quad \text{where} \quad \hat{\pi} = \frac{\sum_i (n_i - 1) S_i}{\sum_i n_i(n_i - 1)} \tag{2.12}$$

- Pseudo Likelihood Estimator

Davidian and Carroll (1987) and Carroll and Ruppert (1988) introduced the pseudo likelihood estimator. Treat the count number $S_i = \sum_i y_{ij}$ as a Gaussian distribution random variable. So the likelihood for S_i is:

$$f(S_i) = \frac{1}{\sqrt{2\pi m_{2i}}} exp\left\{-\frac{1}{2}\frac{(S_i - n_i\pi)^2}{m_{2i}}\right\}$$

Thus the estimating equation is:

$$\begin{aligned} U_{PL} &= \frac{\partial(-2log(f(S_i)) - log(2\pi))}{\partial\rho} \\ &= \frac{\partial\frac{(S_i - n_i\pi)^2}{m_{2i}}}{\partial\rho} + \frac{\partial m_{2i}}{\partial\rho} \\ &= \frac{n_i - 1}{1 + (n_i - 1)\rho}\left[\frac{(S_i - n_i\pi)^2}{m_{2i}} - 1\right] \end{aligned} \tag{2.13}$$

Denote the solution for (2.13) as the pseudo likelihood estimator ρ_{PL}. Note that, ρ_{PL} is different with the Gaussian likelihood estimator ρ_G. ρ_G is got by treating the vector response $y_i = \{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$ as a multivariate normal distribution while ρ_{PL} is got by maxmizing the pseudo likelihood of $S_i = \sum_j y_{ij}$

- Extended Quasi Likelihood Estimator

Nelder and Pregibon (1987) extended the quasi likelihood estimating equation for the common correlation model to estimate the ICC ρ. Note that the traditional quasi likelihood approach can not be used here, since the residuals ε_i does not contain ρ.

The quasi likelihood estimating equation for ρ is:

$$\begin{aligned} U_{EQ} &= \sum_i \left\{ \frac{n_i - 1}{[1 + (n_i - 1)\rho]^2} \left[D_i(S_i, \pi) - [1 + (n_i - 1)\rho] \right] \right\} \\ &= \sum_i (n_i - 1) \left[\frac{D_i(S_i, \pi) - [1 + (n_i - 1)\rho]}{[1 + (n_i - 1)\rho]^2} \right] \end{aligned} \tag{2.14}$$

Here

$$D_i(S_i, \pi) = 2 \left[S_i \ log(\frac{S_i}{n_i \pi}) + (n_i - S_i) \ log(\frac{n_i - S_i}{n_i - n_i \pi}) \right]$$

Denote the solution for (2.14) as the quasi likelihood estimator ρ_Q^*. It is inconsistent since E $D_i(S_i, \pi) \neq 1 + (n_i - 1)\rho$. One way to correct the inconsistency is to replace $D_i(S - i, \pi)$ with $X_i^2 = \dfrac{(S_i - n_i\pi)^2}{n_i \pi (1 - \pi)}$. This will yields the pseudo likelihood estimator ρ_P. Another way is to replace $D_i(S_i, \pi)$ with $\frac{k}{k-1} D_i(S_i, \pi)$, this will yield the unbiased version of the quasi likelihood estimator ρ_{EQ}.

- Moment Estimator

Kleinman (1973) proposed a set of moment estimators in the form of:

$$\hat{\rho}_M = \frac{S_\omega - \tilde{\pi}_\omega(1-\tilde{\pi}_\omega)\sum_i \frac{\omega_i(1-\omega_i)}{n_i}}{\tilde{\pi}_\omega(1-\tilde{\pi}_\omega)\left[\sum_i \omega_i(1-\omega_i) - \sum_i \frac{\omega_i(1-\omega_i)}{n_i}\right]} \tag{2.15}$$

Where $\tilde{\pi}_\omega = \sum_i \omega_i \tilde{\pi}_i$ is the weighted average of $\tilde{\pi}_i = \dfrac{S_i}{n_i}$ and $S_\omega = \sum_i \omega_i(\tilde{\pi}_i - \tilde{\pi}_\omega)^2$ is the weighted variance of $\tilde{\pi}_i$. (2.15) is derived by equating $\tilde{\pi}_\omega$ and S_ω to their expected values under the common correlation model.

Two specifications of the moment estimators are used in Ridout et al. (1999), one with weights ($\omega_i = 1/k$) and the other with ($\omega_i = n_i/N$). They are labeled ρ_{KEQ} and ρ_{KPR}. If S_ω is replaced by $S^*_\omega = \dfrac{k-1}{k}S_\omega$, we can get two slightly different moment estimators ρ^*_{KEQ} and ρ^*_{KPR}.

A more general moment estimator proposed by Whittle (1982) is by using the iterative algorithms. Take $\omega_i = \dfrac{n_i}{1+(n_i-1)\hat{\rho}}$, where $\hat{\rho}$ is the current estimate of ρ, we can get a new moment estimator ρ_W and ρ^*_W (by replacing S_ω with S^*_ω mentioned above).

- Estimators Based on Quadratic Estimating Equations

The quadratic estimating equations was first proposed by Crowder (1987). It is a set of estimating equations with the quadratic form of $S_i - n_i\pi$:

$$\begin{aligned} U_{QEE}(\pi;\rho) &= \sum_i \left[a_{i\pi}\frac{S_i - n_i\pi}{n_i} + b_{i\pi}\frac{(S_i-n_i\pi)^2 - m_{2i}}{n_i^2}\right] \\ U_{QEE}(\rho,\pi) &= \sum_i \left[a_{i\rho}\frac{S_i - n_i\pi}{n_i} + b_{i\rho}\frac{(S_i-n_i\pi)^2 - m_{2i}}{n_i^2}\right] \end{aligned}$$

He also proposed that the optimal estimating equations is obtained by setting:

$$a_{i\pi} = \frac{-(\gamma_{2i\lambda}+2)+\gamma_{1i\lambda}(1-2\pi)\sigma_{i\lambda}/\pi(1-\pi)}{\sigma_{i\lambda}^2\gamma_{i\lambda}},$$
$$b_{i\pi} = \frac{\gamma_{1i\lambda}-(1-2\pi)\sigma_{i\lambda}/\pi(1-\pi)}{\sigma_{i\lambda}^3\gamma_{i\lambda}},$$

and

$$a_{i\rho} = \frac{\gamma_{1i\lambda}\pi(1-\pi)(n_i-1)}{n_i\sigma_{i\lambda}^3\gamma_{i\lambda}},$$
$$b_{i\rho} = \frac{-\pi(1-\pi)(n_i-1)}{n_i\sigma_{i\lambda}^4\gamma_{i\lambda}}$$

Here γ_{1j} and γ_{2j} are the skewness and kurtosis of $\frac{S_i - n_i\pi}{n_i}$ and $\sigma_{i\lambda}$ is the variance of $\frac{S_i - n_i\pi}{n_i}$. However we do not know the exact form of γ_{1i} and γ_{2i} for the non likelihood estimators. Paul (2001) suggested to use the 3^{rd} and 4^{th} moments derived from the beta-binomial distribution instead:

$$\mu_{2i} = \pi(1-\pi)\{1+(n_i-1)\rho\}/n_i,$$

$$\mu_{3i} = \mu_{2i}(1-2\pi)\{1+(2n_i-1)\rho\}/n_i(1+\rho),$$

and

$$\mu_{4i} = \mu_{2i}\frac{1-\rho}{(1+\rho)(1+2\rho)n_i^2}\left[\frac{\{1+(2n_i-1)\rho\}\{1+(3n_i-1)\rho\}\{1-3\pi(1-\pi)\}}{1-\rho} + (n_i-1)\{\rho+3n_i\mu_{2i}\}\right]$$

Denote this estimator as ρ_{QB}. It is supposed to be the optimal quadratic estimating equations for the beta binomial data.

The Gaussian likelihood estimator and pseudo likelihood estimator are special cases of the optimal quadratic estimating equations. For the Gaussian likelihood estimator, the parameters are:

$$a_{i\rho} = n_i(1-2\pi) \quad \text{and} \quad b_{i\rho} = \frac{n_i^2[1+(n_i-1)\rho^2]}{[1+(n_i-1)\rho]^2}$$

For the pseudo likelihood estimator, the parameters are:

$$a_{i\rho} = 0 \quad \text{and} \quad b_{i\rho} = \frac{n_i^2(n_i-1)}{[1+(n_i-1)\rho]m_{2i}}$$

It also coincides with the optimal estimating equations when we set $\gamma_{1i} = \gamma_{2i} = 0$

2.3 The Past Comparisons of the Estimators

Ridout et al. (1999) compared 20 estimators of the intra-class coefficient for their bias, standard deviation, mean square error and relative efficiency. He suggested that the analysis of variance estimator (ρ_A), the κ-type estimator (ρ_{FC}) and the moment estimator (ρ_{KPR} and ρ_W) performed well as far as the median of the mean square error were concerned. He also found that the Pearson estimator ($\rho_{Pearson}$) performed well when the true value of the intra-class correlation parameter ρ was small. But the overall performance of $\rho_{Pearson}$ depends on the true value of ρ. The conclusion of Ridout et al. (1999) were based on the simulation results on the data generated from the beta binomial distribution and the mixed distribution of two binomial distributions.

Paul (2003) introduced 6 other estimators based on the quadratic estimating equations and compare these 6 estimators along with the 20 estimators used by Ridout et al. (1999). With similar setup of the simulation, Paul (2003) showed that the estimator based on the optimal quadratic estimating equations ρ_{QB}, which used the 3^{rd} and 4^{th} moment from beta binomial distribution, also performs well in the jointly estimation of $(\hat{\pi}, \hat{\rho})$. For the data generated from the beta binomial distribution, it even has higher efficiency than that of ρ_A. He also found that the performance of $\rho_{Pearson}$ depends on the true value of ρ, which is consistent with the findings of Ridout et al.(1999).

Zou and Donner (2004) introduced the coverage rate as a new index to compare the performance of the estimators. They obtained the closed form of the asymptotic variances of the analysis of variance estimator ρ_A, the κ-type estimator ρ_{FC} and the Pearson estimator $\rho_{Pearson}$, under the distribution of the generalized binomial models (Madsen, 1993). The simulation results indicated that the κ type estimator ρ_{FC} performed best among these three estimators as far as the coverage rate of the confidence interval was concerned.

2.4 The Estimators We Compare

We are going to compare four estimators. The κ-type estimator ρ_{FC}, the analysis of variance estimator ρ_A, the Gaussian estimator ρ_G and the UJ estimator based on the Cholesky decomposition.

The κ-type estimator ρ_{FC} and the ANOVA estimator ρ_A are widely used estimators for ICC and performs well in many situations (Ridout et al. 1999). Gaussian likelihood method is the most general form of all the moment based methods. And it also only relies on the first two moments, that is what we know in the common correlated model. Besides, the Gaussian likelihood method is also the most convenient to calculate method of all the pseudo likelihood methods. (Crowder 1985).

We are going to compare these three estimators with the new estimator ρ_{UJ} based on the Cholesky decomposition, which is the specification of the UJ method proposed by Wang and Carey (2004).

2.5 The Properties of the Estimators

2.5.1 The Asymptotic Variances of the Estimators

The asymptotic variance quantifies the limit properties of the estimators. As shown above, we have two types of estimators for ρ. One type of the estimator is the solution of some estimating equation, such as the new estimator ρ_{UJ} and the Gaussian Likelihood estimator ρ_G. Another type of the estimator has the closed form, such as the κ-type estimator ρ_{FC} and the Anova estimator ρ_A. We may use different methods to obtain the asymptotic variances of these two types of estimators.

- Estimators Without Closed Forms

  This type of the estimator is the solution of some estimating equation and has no closed forms. The typical example is the NEW (UJ) estimator.

$$U_J(\rho;\pi) = \sum_i \frac{C_i}{n_i(n_i-1)} \left[(1-2\pi)[1+(n_i-1)\rho](S_i - n_i\pi) - (S_i - n_i\pi)^2 + m_{2i}\right]$$

  Note that the π in the estimating equation is also unknown and we need to solve the estimating equations for (π, ρ) jointly. So the choice of the estimators of $\hat{\pi}$ may affect the asymptotic variance of $\hat{\rho}$. Here we will use (2.2):

$$U(\pi; \rho) = \sum_{i=1}^{k} \frac{S_i - n_i\pi}{1 + (n_i - 1)\rho}$$

as the estimating equation for $\hat{\pi}$. The advantage of this estimator is that it would maximize the efficiency of $\hat{\pi}$.

Of all the estimators mentioned above, the MLE estimator(ρ_{ML}), the Gaussian estimator (ρ_G), the Pseudo likelihood estimator (ρ_{PL}), the extended quasi-likelihood estimator (ρ_{EQ}), the estimator based on the quadratic estimating equations (ρ_{QB}) and the New (UJ) estimator ρ_{UJ} based on Cholesky decomposition are of this type.

For this type of estimators, we shall use the sandwich method to get the asymptotic variance-covariance matrix of $(\hat{\pi}, \hat{\rho})$. Assume the joint estimating equations for $\theta = (\pi, \rho)$ are:

$$U(\theta) = \begin{pmatrix} U(\pi; \rho) \\ U(\rho; \pi) \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \tilde{g}$$

and define

$$\begin{aligned} \Delta &= -\mathrm{E}(\partial \tilde{g} / \partial \theta^T) \\ &= \begin{pmatrix} -\mathrm{E}\dfrac{\partial g_1}{\partial \pi} & -\mathrm{E}\dfrac{\partial g_1}{\partial \rho} \\ -\mathrm{E}\dfrac{\partial g_2}{\partial \pi} & -\mathrm{E}\dfrac{\partial g_2}{\partial \rho} \end{pmatrix} = \begin{pmatrix} d_1 & d_4 \\ d_2 & d_3 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} M &= \mathrm{Var}(\tilde{g}) \\ &= \begin{pmatrix} \mathrm{Var}(g_1) & \mathrm{Cov}(g_1, g_2) \\ \mathrm{Cov}(g_2, g_1) & \mathrm{Var}(g_2) \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \end{aligned}$$

So the asymptotic variance-covariance matrix is

$$\mathrm{Var}(\hat{\pi},\hat{\rho}) = \Delta^{-1}M(\Delta^T)^{-1} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

Here $\mathrm{Var}(\hat{\rho}) = V_{22}$. Simply plugging in the estimates of $(\hat{\pi},\hat{\rho})$ can not ensure the positiveness of matrix M and sometimes we will get the negative values of the asymptotic variances of $\hat{\rho}_G$ and $\hat{\rho}_{UJ}$. Here we define:

$$M^{\sharp} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} \sum(g_{1i}^2) & \sum g_{1i}g_{2i} \\ \sum g_{1i}g_{2i} & \sum(g_{1i}^2) \end{pmatrix}$$

$M^{\sharp}$ is a positive matrix. So, use $M^{\sharp}$ instead of M if necessary, then the asymptotic variance of $\hat{\rho}_G$ and $\hat{\rho}_{UJ}$ will always be positive.

For our choice of the estimating equation for π:

$$g_1 = \sum_i \frac{S_i - n_i\pi}{1+(n_i-1)\rho}$$

We have:

$$d_1 = -\mathrm{E}\frac{\partial g_1}{\partial \pi} = \sum_i \frac{n_i}{1+(n_i-1)\rho}$$

and

$$d_4 = -\mathrm{E}\frac{\partial g_1}{\partial \rho} = -\mathrm{E}(S_i - n_i\pi)(-\frac{n_i-1}{[1+(n_i-1)\rho]^2}) = 0$$

So

$$\Delta^{-1} = \begin{pmatrix} \frac{1}{d_1} & 0 \\ -\frac{d_2}{d_1 d_3} & \frac{1}{d_3} \end{pmatrix}$$

And for the M, we have:

$$M_{11} = \mathrm{Var}(g_1) = \mathrm{E}\sum_i \frac{(S_i - n_i\pi)^2}{[1+(n_i-1)\rho]^2} = \sum_i \frac{n_i\pi(1-\pi)}{1+(n_i-1)\rho}$$

Thus:

$$\begin{aligned}\mathrm{Var}(\hat{\pi}) = V_{11} &= \left(\frac{1}{d_1}, 0\right)\mathrm{Var}(\tilde{g})\left(\frac{1}{d_1}, 0\right)^T \\ &= \frac{M_{11}}{d_1^2} = \left(\sum_i \frac{n_i^2}{m_{2i}}\right)^{-1} \end{aligned} \quad (2.16)$$

$$\begin{aligned}\mathrm{Var}(\hat{\rho}) = V_{22} &= \left(-\frac{d_2}{d_3 d_1}, \frac{1}{d_3}\right)\mathrm{Var}(\tilde{g})\left(-\frac{d_2}{d_3 d_1}, \frac{1}{d_3}\right)^T \\ &= \frac{1}{d_3^2}\left[M_{22} - 2M_{12}\frac{d_2}{d_1} + (\frac{d_2}{d_1})^2 M_{11}\right] \end{aligned} \quad (2.17)$$

where $m_{2i} = \mathrm{E}(S_i - n_i\pi)^2 = n_i\pi(1-\pi)[1+(n_i-1)\rho]$ is the 2^{nd} order centralized moment of S_i. m_{3i} and m_{4i} are the 3^{rd} and 4^{th} order centralized moment of S_i

Apply the sandwich method on the NEW(UJ) estimator and the Gaussian likelihood estimator, with m_{3i} and m_{4i} to denote the 3^{rd} and 4^{th} order centralized moment of S_i.

For the NEW(UJ) estimator, the estimating equation is:

$$g_2 = U_J = \sum_i \frac{C_i}{n_i(n_i-1)}\{(1-2\pi)[1+(n_i-1)\rho](S_i - n_i\pi) - [(S_i - n_i\pi)^2 - m_{2i}]\}$$

thus we have:

$$\begin{aligned} d_2 &= -\mathrm{E}\frac{\partial g_2}{\partial \pi} \\ &= \sum_i \frac{1}{n_i(n_1-1)}\frac{\partial C_i}{\partial \pi}\mathrm{E}\left\{(1-2\pi)[1+(n_i-1)\rho](S_i - n_i\pi) - \left[(S_i - n_i\pi)^2 - m_{2i}\right]\right\} \\ &\quad + \sum_i \frac{C_i}{n_i(n_i-1)}\mathrm{E}\frac{\partial\{(1-2\pi)[1+(n_i-1)\rho](S_i - n_i\pi) - [(S_i - n_i\pi)^2 - m_{2i}]\}}{\partial \pi} \\ &= 0 + \sum_i \frac{C_i}{n_i(n_i-1)}\mathrm{E}\Big\{ -2[1+(n_i-1)\rho](S_i - n_i\pi) - n_i(1-2\pi)[1+(n_i-1)\rho] \\ &\quad -2(S_i - n_i\pi)(-n_i) + n_i(1-2\pi)[1+(n_i-1)\rho]\Big\} \\ &= 0 \end{aligned} \quad (2.18)$$

$$
\begin{aligned}
d_3 &= -\mathrm{E}\frac{\partial g_2}{\partial \rho} \\
&= \mathrm{E}\sum_i \frac{\partial \frac{C_i}{n_i(n_i-1)}}{\partial \rho}\{(1-2\pi)[1+(n_i-1)\rho](S_i-n_i\pi)-[(S_i-n_i\pi)^2-m_{2i}]\} \\
&\quad + \frac{C_i}{n_i(n_i-1)}\frac{\partial\{(1-2\pi)[1+(n_i-1)\rho](S_i-n_i\pi)-[(S_i-n_i\pi)^2-m_{2i}]\}}{\partial \rho} \\
&= \mathrm{E}\sum_i \left(0+\frac{C_i}{n_i(n_i-1)}\Big\{(1-2\pi)[1+(n_i-1)\rho](S_i-n_i\pi)\right. \\
&\quad \left. -[0-n_i\pi(1-\pi)[1+(n_i-1)\rho]]\Big\}\right) \\
&= \sum_i C_i\pi(1-\pi) \qquad (2.19)
\end{aligned}
$$

and

$$
\begin{aligned}
M_{22} &= \sum_i \frac{C_i^2}{n_i^2(n_i-1)^2}\times\{m_{4i}-2(1-2\pi)[1+(n_i-1)\rho]m_{3i} \\
&\quad +(1-2\pi)^2[1+(n_i-1)\rho]^2 m_{2i}-m_{2i}^2\} \qquad (2.20)
\end{aligned}
$$

So the variance for $\hat{\rho}_{UJ}$ is:

$$
\begin{aligned}
\mathrm{Var}(\hat{\rho}_{UJ}) &= \frac{1}{d_3^2}\left[M_{22}-2M_{12}\frac{d_2}{d_1}+(\frac{d_2}{d_1})^2 M_{11}\right]=\frac{M_{22}}{d_3^2} \\
&= \sum_i \frac{C_i^2}{[n_i(n_i-1)\pi(1-\pi)]^2(\sum_i C_i)^2}\times\Big\{m_{4i}-2(1-2\pi)[1+(n_i-1)\rho]m_{3i} \\
&\quad +(1-2\pi)^2[1+(n_i-1)\rho]^2 m_{2i}-m_{2i}^2\Big\} \qquad (2.21)
\end{aligned}
$$

We can see that since $d_2=-\mathrm{E}\frac{\partial g_2}{\partial \pi}=0$, $\mathrm{Var}(\hat{\rho}_{UJ})$ does not depend on the choice of g_1. This is not always true since d_2 may not equal to 0 for other estimators, such as the Gaussian likelihood estimator. However, $d_4=-\mathrm{E}\frac{\partial g_1}{\partial \rho}=0$ is always true for our choice of the estimating equation for π, so $\mathrm{Var}(\hat{\pi})$ does not depend on the choice of $\hat{\rho}$, which is always $(\sum_i n_i^2/m_{2i})^{-1}$

For the Gaussian likelihood estimator, the estimating equation is:

$$g_2 = U_G = \sum_i \left\{(1-2\pi)(S_i - n_i\pi) - \frac{1+(n_i-1)\rho^2}{[1+(n_i-1)\rho]^2}[(S_i-n_i\pi)^2 - m_{2i}]\right\}$$

thus we have:

$$\begin{aligned} d_2 = -\mathrm{E}\frac{\partial g_2}{\partial \pi} &= \sum_i \left\{n_i(1-2\pi) - \frac{1+(n_i-1)\rho^2}{1+(n_i-1)\rho} n_i(1-2\pi)\right\} \\ &= \sum_i \frac{n_i(n_i-1)\rho(1-\rho)(1-2\pi)}{1+(n_i-1)\rho} \end{aligned} \tag{2.22}$$

$$\begin{aligned} d_3 = -\mathrm{E}\frac{\partial g_2}{\partial \rho} &= 0 + \mathrm{E}\sum_i \frac{\partial \frac{1+(n_i-1)\rho^2}{[1+(n_i-1)\rho]^2}[(S_i-n_i\pi)^2 - m_{2i}]}{\partial \rho} \\ &= -\frac{n_i(n_i-1)[1+(n_i-1)\rho^2]\pi(1-\pi)}{[1+(n_i-1)\rho]^2} \end{aligned} \tag{2.23}$$

and:

$$M_{11} = \mathrm{Var}(g_1) = \mathrm{E}\sum_i \frac{(S_i-n_i\pi)^2}{[1+(n_i-1)\rho]^2} = \sum_i \frac{n_i\pi(1-\pi)}{1+(n_i-1)\rho}$$

$$M_{12} = M_{21} = \sum_i \left[\frac{1-2\pi}{1+(n_i-1)\rho} m_{2i} - \frac{1+(n_i-1)\rho^2}{[1+(n_i-1)\rho]^3} m_{3i}\right] \tag{2.24}$$

$$\begin{aligned} M_{22} = \sum_i \Bigg\{&(1-2\pi)^2 m_{2i} + \left[\frac{1+(n_i-1)\rho^2}{[1+(n_i-1)\rho]^2}\right]^2 (m_{4i} - m_{2i}^2) \\ &- \frac{2(1-2\pi)[1+(n_i-1)\rho^2]}{[1+(n_i-1)\rho]^2} m_{3i}\Bigg\} \end{aligned} \tag{2.25}$$

Take these values into (2.17):

$$\begin{aligned} \mathrm{Var}(\hat\rho_G) = V_{22} &= \left(-\frac{d_2}{d_3 d_1}, \frac{1}{d_3}\right) \mathrm{Var}(\tilde g) \left(-\frac{d_2}{d_3 d_1}, \frac{1}{d_3}\right)^T \\ &= \frac{1}{d_3^2}\left[M_{22} - 2M_{12}\frac{d_2}{d_1} + (\frac{d_2}{d_1})^2 M_{11}\right] \end{aligned} \tag{2.26}$$

- Estimators With Closed Forms

Another type of the estimator is the estimator that has closed forms, such as the κ type estimator (2.9):

$$\rho_{FC} = 1 - \frac{1}{(N-k)\hat{\pi}(1-\hat{\pi})} \sum_i \frac{S_i(n_i - S_i)}{n_i}, \qquad \hat{\pi} = \frac{\sum_i S_i}{\sum_i n_i}$$

The $\hat{\pi}$ in (2.9) has been defined clearly as $\hat{\pi} = \sum_i S_i / \sum_i n_i$ in (2.9). So $\hat{\rho}_{FC}$ is a function of (S_i, n_i, k).

Of all the estimators we have mentioned in the last section, the moment estimator $(\rho_W, \rho_{KEQ}, \rho_{KPR})$, the analysis of variance estimator ρ_A, the direct probability interpretation estimator (ρ_{FC}, ρ_{MAK}) and the direct calculation of correlation estimator $\rho_{Pearson}$ are of this type.

We may choose appropriate functions as the intermediate variables and then apply the delta method to obtain the asymptotic variance for these estimators.

Define $Y1 = \sum S_i$ and $Y2 = \sum S_i^2/n_i$. So the covariance-variance matrix of $(Y1, Y2)$ is:

$$\Sigma = \begin{pmatrix} \mathrm{Var}(Y1) & \mathrm{Cov}(Y1, Y2) \\ \mathrm{Cov}(Y2, Y1) & \mathrm{Var}(Y2) \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} \mathrm{Var}(S_i) & \mathrm{Cov}(S_i, S_i^2/n_i) \\ \mathrm{Cov}(S_i^2/n_i, S_i) & \mathrm{Var}(S_i^2/n_i) \end{pmatrix}$$

Define the derivative of $\hat{\rho}$ on $(Y1, Y2)$ as Φ:

$$\Phi = \begin{pmatrix} \frac{\partial \hat{\rho}}{\partial Y1} \\ \frac{\partial \hat{\rho}}{\partial Y2} \end{pmatrix}$$

Application of the Delta method(Agresti, 2002, p.579) yields the asymptotic distribution for $\hat{\rho}$ as:

$$\hat{\rho} - \rho \sim N(0, \Phi^T \Sigma \Phi)$$

So

$$\text{Var}(\hat{\rho}) = (\frac{\partial \hat{\rho}}{\partial Y1})^2 \text{Var}(Y1) + 2(\frac{\partial \hat{\rho}}{\partial Y1}\frac{\partial \hat{\rho}}{\partial Y2})\text{Cov}(Y1, Y2) + (\frac{\partial \hat{\rho}}{\partial Y2})^2 \text{Var}(Y2)$$

which is evaluated at

$$Y1 = \text{E}Y1 = N\pi \quad , \quad Y2 = \text{E}Y2 = \pi(1-\pi)(k + (N-k)\rho) + N\pi^2$$

Similar with estimators without closed form, simply plugging in the estimates of $(\hat{\pi}, \hat{\rho})$ can not ensure the positiveness of Σ, sometimes we will get negative values of the asymptotic variance of ρ_{FC} and $\hat{\rho}_A$. Here we define:

$$\Sigma^\sharp = \begin{pmatrix} \sum(S_i - \overline{S}_i)^2 & \sum(S_i - \overline{S_i})(S_i^2/ni - \overline{S_i^2/ni}) \\ \sum(S_i - \overline{S_i})(S_i^2/ni - \overline{S_i^2/ni}) & \sum(S_i^2/ni - \overline{S_i^2/ni})^2 \end{pmatrix}$$

$\Sigma^\sharp$ is a positive definite matrix. So use $\Sigma^\sharp$ instead of Σ if necessary, we can always get the positive asymptotic variance of $\hat{\rho}_{FC}$ and $\hat{\rho}_A$

Use nm_{li} to denote the l^{th} order noncentralized moment $(\text{E}S_i^l)$ and m_{li} to denote the l^{th} order centralized moment $(\text{E}(S_i - n_i\pi)^l)$. So Σ is:

$$\Sigma = \begin{pmatrix} \sum(nm_{2i} - nm_{1i}^2) & \sum \frac{1}{n_i}(nm_{3i} - nm_{1i}nm_{2i}) \\ \sum \frac{1}{n_i}(nm_{3i} - nm_{1i}nm_{2i}) & \sum \frac{1}{n_i^2}(nm_{4i} - nm_{2i}^2) \end{pmatrix}$$

or

$$\Sigma = \begin{pmatrix} \sum m_{2i} & \sum (m_{3i}/n_i + 2\pi m_{2i}) \\ \sum (m_{3i}/n_i + 2\pi m_{2i}) & \sum \frac{1}{n_i^2} \left[m_{4i} + 4m_{3i}(n_i\pi) + 4m_{2i}(n_i\pi)^2 - m_{2i}^2 \right] \end{pmatrix}$$

Apply the Delta method on the κ-type estimator ρ_{FC} and the anova estimator ρ_A with m_{3i} and m_{4i} to denote the 3^{rd} and 4^{th} order centralized moment.

The κ type estimator is by definition

$$\hat{\rho}_{FC} = 1 - \frac{\sum S_i(n_i - S_i)/n_i}{(N-k)\hat{\pi}(1-\hat{\pi})} \quad where \quad \hat{\pi} = \frac{\sum S_i}{\sum n_i} = \frac{Y1}{N}$$

So,

$$\hat{\rho}_{FC} = 1 - \frac{N^2}{N-k} \frac{Y1 - Y2}{Y1(N - Y1)}$$

Thus the derivatives of $\hat{\rho}_{FC}$ are:

$$\frac{\partial \hat{\rho}_{FC}}{\partial Y1} = \frac{2(N-k)(1-\rho)\pi + N\rho + k(1-\rho)}{(N-k)N\pi(1-\pi)} \quad and \quad \frac{\partial \hat{\rho}_{FC}}{\partial Y2} = -\frac{1}{(N-k)\pi(1-\pi)}$$

Take these values into (2.27) and substitute $(Y1, Y2)$ with $(\mathrm{E}Y1, \mathrm{E}Y2)$, then the asymptotic variance for $\hat{\rho}_{FC}$ is:

$$\begin{aligned} \mathrm{Var}(\hat{\rho}_{FC}) &= \sum \Big\{ \frac{N^2}{n_i^2} m_{4i} - \frac{2N\left[N\rho + k(1-\rho)\right](1-2\pi)}{n_i} m_{3i} \\ &\quad + \left[N\rho + k(1-\rho)\right]^2 (1-2\pi)^2 m_{2i} - \frac{N^2}{n_i^2}(m_{2i})^2 \Big\} \\ &\quad / \left[N^2(N-k)^2\pi^2(1-\pi)^2\right] \end{aligned} \tag{2.27}$$

The ANOVA estimator is by definition:

$$\hat{\rho}_A = \frac{MSB - MSW}{MSB + (n_A - 1)MSW}$$

where

$$MSB = \frac{1}{k-1}\left\{\sum \frac{Y_i^2}{n_i} - \frac{(\sum Y_i)^2}{N}\right\} \quad , \quad MSW = \frac{1}{N-k}\left\{\sum Y_i - \sum \frac{Y_i^2}{n_i}\right\}$$

and

$$n_A = \frac{1}{k-1}\left(N - \frac{\sum n_i^2}{N}\right)$$

So

$$\hat{\rho}_A = \frac{Y1[kS1 - N(Y1 + k - 1)] + Y2N(N-1)}{Y1[N(k-1)(n_A - 1) - Y1(N-k)] + Y2N[N - 1 - nA(k-1)]}$$

Thus the derivatives of $\hat{\rho}_A$ are:

$$\frac{\partial \hat{\rho}_A}{\partial Y1} = -\frac{(k-1)n_A\,[k(1-2\pi)(1-\rho) + N\,(2\pi(1-\rho) + \rho)]}{(N-k)\pi(1-\pi)[1 + (k-1)(1-\rho)n_A + \rho(N-1)]^2}$$

$$\frac{\partial \hat{\rho}_A}{\partial Y2} = \frac{(k-1)n_A N}{(N-k)\pi(1-\pi)[1 + (k-1)(1-\rho)n_A + \rho(N-1)]^2}$$

Similar with the calculation for $\hat{\rho}_{FC}$, the asymptotic variance of $\hat{\rho}_A$ is:

$$\begin{aligned} \text{Var}(\hat{\rho}_A) \;=\; & \sum \Big\{ \frac{(k-1)^2 N^2 n_A^2}{n_i^2} m_{4i} - \frac{2(k-1)^2 N n_A^2 (1-2\pi)\,[\rho N + k(1-\rho)]}{n_i} m_{3i} \\ & + (k-1)^2 n_A^2 (1-2\pi)^2\,[\rho N + k(1-\rho)]^2\, m_{2i} - \frac{(k-1)^2 N^2 n_A^2}{n_i^2} m_{2i}^2 \Big\} \\ & / \left\{(N-k)^2 \pi^2 (1-\pi)^2 [1 + (k-1)n_A(1-\rho) + (N-1)\rho]^4\right\} \qquad (2.28) \end{aligned}$$

As mentioned before, we can get the closed form of 3^{rd} and 4^{th} order centralized moment m_{3i} and m_{4i} in the parametric model.

For the Generalized Binomial model,we have

$$\begin{aligned}
\mathrm{Var}(Y1) &= \pi(1-\pi)\sum n_i\pi(1-\pi)[1+(n_i-1)\rho] = \pi(1-\pi)\left[\rho\sum n_i^2+(1-\rho)N\right] \\
\mathrm{Cov}(Y1,Y2) &= \mathrm{Cov}(Y2,Y1) = \pi(1-\pi)\left[\rho\sum n_i^2+2\pi(1-\rho)N+(1-\rho)(1-2\pi)k\right] \\
\mathrm{Var}(Y2) &= \pi(1-\pi)\Big\{[1-6\pi+6\pi^2](1-\rho)\sum\frac{1}{n_i}+\pi[6+\rho-\pi(10+\rho)](1-\rho)k \\
&\quad +2\pi[-\rho+\pi(2+\rho)](1-\rho)N+\rho[1+\pi(1-\pi)(1-\rho)]\sum n_i^2\Big\}
\end{aligned}$$

So the variance of $\hat{\rho}_{FC}$ under the Generalized Binomial distribution is:

$$\begin{aligned}
\mathrm{Var}(\hat{\rho}_{FC}) &= -(1-\rho)/\left[(N-k)^2N^2\pi(1-\pi)\right] \\
&\quad \times\left\{N\left[k^2(1-2\pi)^2-2kN\pi(1-\pi)+N(-1+6\pi(1-\pi))\sum\frac{1}{n_i}\right]\right. \\
&\quad -\rho\Big[k^2(1-2\pi)^2\sum n_i^2+N^2(\sum n_i^2-\pi(1-\pi)(2N+3\sum n_i^2)) \\
&\quad +Nk(-2\sum n_i^2+\pi(1-\pi)(N+8\sum n_i^2))\Big] \\
&\quad \left.-(N-k)^2(1-2\pi)^2(N-\sum n_i^2)\rho^2\right\} \qquad (2.29)
\end{aligned}$$

Or, in form of the cubic function of ρ,it is:

$$\begin{aligned}
\mathrm{Var}(\hat{\rho}_{FC}) &= (1-\rho) \\
&\quad \times\left\{\left[\frac{1}{\pi(1-\pi)}-6\right]\frac{\sum\frac{1}{n_i}}{(N-k)^2}\right. \\
&\quad +\left[2N+4k-\frac{k}{\pi(1-\pi)}\right]\frac{k}{N(N-k)^2} \\
&\quad +\Big[\frac{\sum n_i^2}{N^2\pi(1-\pi)} \\
&\quad -\frac{(3N-2k)(N-2k)\sum n_i^2}{N^2(N-k)^2}-\frac{2N-k}{(N-k)^2}\Big]\rho \\
&\quad \left.+\left[4-\frac{1}{\pi(1-\pi)}\right]\frac{\sum n_i^2-N}{N^2}\rho^2\right\}
\end{aligned}$$

For the analysis of variance estimator, the asymptotic variance of ρ_A is:

$$\begin{aligned}
\mathrm{Var}(\hat{\rho}_A) &= \frac{(k-1)^2 n_A^2 (1-\rho)}{\{(N-k)^2\pi(1-\pi)\left[1+(k-1)(1-\rho)n_A+\rho(N-1)\right]^4\}} \times \\
&\Bigg\{ -k^2(1-2\pi)^2(1-\rho)(N+N\rho-\sum n_i^2\rho)+ \\
&kN\Big[2N\pi(1-\pi)-2\rho\sum n_i^2+\rho\pi(1-\pi)(N+8\sum n_i^2) \\
&-2\rho^2(1-2\pi)^2(N-\sum n_i^2)\Big]+ \\
&N^2\Big[(1-6\pi+6\pi^2)\sum\frac{1}{n_i}+\rho\sum n_i^2 \\
&+\rho\left(-\pi(1-\pi)(2N+3\sum n_i^2)+\rho(1-2\pi)^2(N-\sum n_i^2)\right)\Big] \\
&\Bigg\}
\end{aligned}$$

A more easy to read form is:

$$\begin{aligned}
\mathrm{Var}(\hat{\rho}_A) &= \frac{(k-1)^2 n_A^2 (1-\rho)}{(N-k)^2\left[1+(k-1)n_A(1-\rho)+\rho(N-1)\right]^4} \times \\
\Bigg\{ &N\left[2Nk+k^2\left(4-\frac{1}{\pi(1-\pi)}\right)+N\left(\frac{1}{\pi(1-\pi)}-6\right)\sum\frac{1}{n_i}\right] \\
&+\rho\Big\{-N^2(2N-k)+ \\
&\left[k(2N-k)\left(4-\frac{1}{\pi(1-\pi)}\right)+N^2\left(\frac{1}{\pi(1-\pi)}-3\right)\right]\sum n_i^2\Big\} \\
&+\rho^2\left[(N-k)^2(N-\sum n_i^2)\left(\frac{1}{\pi(1-\pi)}-4\right)\right]\Bigg\}
\end{aligned}$$

2.5.2 The Relationship of the Asymptotic Variances

- Note that:

$$\mathrm{Var}(\hat{\rho}_A)/\mathrm{Var}(\hat{\rho}_{FC}) = \frac{N^2 n_A^2 (k-1)^2}{[1+(k-1)n_A(1-\rho)+(N-1)\rho]^4}$$

When ρ takes the extreme value 1, it could be reduced to $(N - \sum n_i^2/N)^2/N^2$, which is between $(0, (\frac{k-1}{k})^2)$. That means when ρ is large enough, the variance of the ANOVA estimator is smaller that that of the FC estimator.

- In the balanced design ($n_1 = n_2 = \ldots = N/k$), $\mathrm{Var}(\hat{\rho}_{UJ})$ and $\mathrm{Var}(\hat{\rho}_{FC})$ will converge to the same value:

$$\begin{aligned} \mathrm{Var}(\rho) &= \sum_i \Big\{ k^2(m_{4i} - m_{2i}^2) - 2(1-2\pi)k[N\rho + k(1-\rho)]m_{3i} + \\ &\quad (1-2\pi)^2[N\rho + k(1-\rho)]^2 m_{2i} \Big\} \end{aligned} \tag{2.30}$$

Since n_i is constant now, the moment m_{2i}, m_{3i} and m_{4i} are independent of i.

Chapter 3

Simulation Study

3.1 Setup

Simulations were run for four values of mean parameter π $(0.05, 0.1, 0.2, 0.5)$, five values of intra-class correlation parameter ρ $(0.05, 0.1, 0.2, 0.5, 0.8)$, three values of the number of clusters(sample size) k $(10, 25, 50)$, two distributions of the number of individuals within the cluster(cluster size) n_i and three probability distributions of S_i (the summation of the binary responses within the cluster). A fully combination of these five factors are used, giving a total of 360 runs. For each run, we generate 1000 samples.

Note that $P(y_{ij} = 1) = \pi$ and $P(y_{ij}) = 0 = 1 - \pi$ are complement and $\text{Corr}(1 - y_{ij}, 1 - y_{ik}) = \rho$, we do not need to investigate the values of π that is larger than 0.5.

The values from 0.05 to 0.8 of ρ are used to simulate the situations that the response are from almost independent to highly correlated.

The three values of k are used to simulate the small sample size ($k = 10$), the medium sample size($k = 25$) and the large sample size($k = 50$).

The first distribution of the cluster size is in view of the widespread use of the common-correlation model in toxicology study. It was an empirical distributions of 523 litter size. The litter size ranges from 1 to 19, with a mean of 12.0 and standard deviation 2.98. This distribution of cluster size n_i was first quoted by Kupper et.al (1986) and we use "Kupper" to index it.

The second distribution of the cluster size is a truncated negative binomial distribution, ranging from 1 to 15. This distribution is based on the human sibship data for the U.S. (Brass, 1958) with mean 3.1 and standard deviation 2.11. We use Brass to index this distribution of n_i in the thesis.

Table (3.1) is the frequency table of the two distributions of the cluster size n_i. Figure (3.1) shows the difference of the two distributions of cluster size n_i. The mean of the "Brass" distribution of n_i is smaller that that of the "Kupper" distribution of n_i. For the Brass distribution of n_i, the probability that $n_i > 7$ is very small. But for the Kupper distribution, the probability that $n_i < 7$ is very small. In addition, the Brass distribution is left skewed while the Kupper distribution is somewhat symmetric.

Three probability distributions are used to simulate the data with the parameter given above. The first is the beta binomial distribution, the second is the generalized binomial distribution. The third probability distribution is sampled by thresholding a multivariate normal distribution. The procedures are as following:

1. n_i continuous data $\{x_{i1}, x_{i2}, \ldots, x_{in_i}\}$ are sampled from the multivariate normal

Table 3.1: Distributions of the Cluster Size

n_{ij}	Kupper	Brass
1	0.0038	0.17708
2	0.0057	0.21811
3	0.0076	0.20161
4	0.0172	0.15538
5	0.0153	0.10543
6	0.0115	0.06506
7	0.0191	0.03729
8	0.0382	0.02014
9	0.0364	0.01037
10	0.0727	0.00512
11	0.1224	0.00245
12	0.1568	0.00113
13	0.1778	0.00051
14	0.1396	0.00023
15	0.1109	0.0001
16	0.0364	0
17	0.0229	0
18	0.0019	0
19	0.0038	0
mean	11.9816	3.1
s.d	2.98	2.11

distribution with mean vector $\tilde{\mu}$ and variance-covariance matrix $V_i = A_i^{1/2} R_i A_i^{1/2}$. Here $A_i = diag\{\sigma^2\}$ and R_i is the compound symmetry correlation matrix with correlation parameter ρ.

2. define $y_{ij} = I_{\{x_{ij}>0\}}$. We can choose appropriate $\tilde{\mu}$ and σ^2 such that $E(y_{ij}) = \pi$ and $\text{Corr}(y_{ij}, y_{ik}) = \rho$. It can be shown that such $(\tilde{\mu}, \sigma^2)$ always exist.

3. y_{ij} is the common correlated binary data that satisfy $\text{Corr}(y_{ij}, y_{ik}) = \rho$ and $\text{Corr}(y_{ij}, y_{lk}) = 0$. Let $S_i = \sum_{j=1}^{n_i} y_{ij}$, then $\{(n_i, S_i), i = 1, 2, \ldots, k\}$ is the data

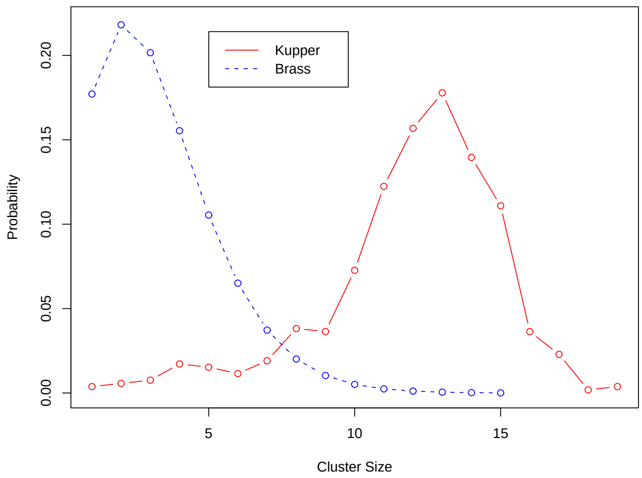

Figure 3.1: The two distributions of the cluster size n_i

that we can estimate.

Note that such kind of data will be rejected:

- $S_i = 0$, for i=1,2,...,k;
- $S_i = n_i$, for i=1,2,...,k;
- $n_i = 1$, for i=1,2,...,k;

It is reasonable to reject these kinds of data since we are not going to estimate them in practice.

1000 acceptable data sets are generated for each combination of the parameters. Then we use the four estimators to estimate the parameter (π, ρ) for each data set. For each estimator and each combination of the parameters, we calculate the bias (Bias=$\sum_i(\hat{\rho}_i - \rho)/1000$), the Standard Deviation(SD=$\sum_i(\hat{\rho}_i - \rho)^2/(1000-1)$), and the mean square error (MSE=$\sum_i(\hat{\rho}_i - \rho)^2/1000$) of the 1000 estimates we get. We also calculate the relative efficiencies of the estimators, using FC estimator as the baseline. It is defined as:

$$R.E = \frac{MSE\{\hat{\rho}_{FC}\}}{MSE\{\hat{\rho}\}} \tag{3.1}$$

3.2 Results

3.2.1 The Overall Performance

First of all, we shall compare the overall performances of the estimators for π and ρ across all the parameter combinations. For π_G and π_{UJ}, they were obtained when we estimate (π, ρ) by solving the estimating equations simultaneously. And for π_A and π_{FC}, they are defined by substituting the $\hat{\rho}$ in (3.2):

$$\hat{\pi} = \sum_i \frac{\frac{S_i}{1+(n_i-1)\hat{\rho}}}{\frac{n_i}{1+(n_i-1)\hat{\rho}}} \tag{3.2}$$

with ρ_A and ρ_{FC}, respectively. Note that (3.2) is equivalent with the estimating equation that we used to estimate π.

$$U(\pi; \rho) = \sum_i \frac{S_i - n_i\pi}{1 + (n_i - 1)\rho}$$

Figure (3.2) is the Box-and-Whisker plot which summarized the bias, standard deviation (SD) and the mean square error (MSE) of the four estimators of ρ (upper row) and π (lower row) when the sample size $k = 10$. The upper row shows the performance of $\hat{\rho}$ and the lower row shows the performance of π. The lower and upper bound of the rectangles in the plot are the 25% and 75% quantiles and the black horizontal line is the median. Figure (3.3) and (3.4) are the Box-and-Whisker plot when $k = 25$ and $k = 50$.

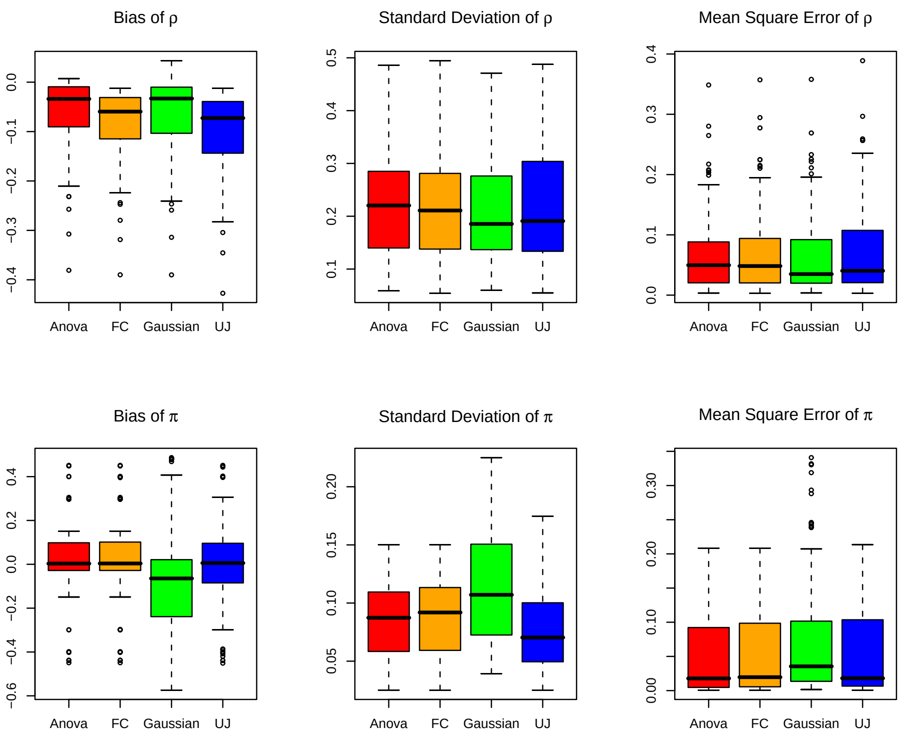

Figure 3.2: The overall performances of the four estimators when $k = 10$.

From these plots, we can see that: for all the four estimators of π, the mean square error and the standard deviation are both very small. As far as the bias is concerned, the estimators π_{FC},π_A and π_{UJ} are nearly unbiased but the Gaussian estimator π_G is negatively skewed. In addition, there exist outliers for the two closed form estimators π_{FC} and π_A but no outliers for π_G and π_{UJ}.

All the four estimators of ρ are negatively skewed. The median of the bias of ρ_A is the smallest while that of ρ_{UJ} is the largest. The 25% quantile of ρ_{UJ} is lower than

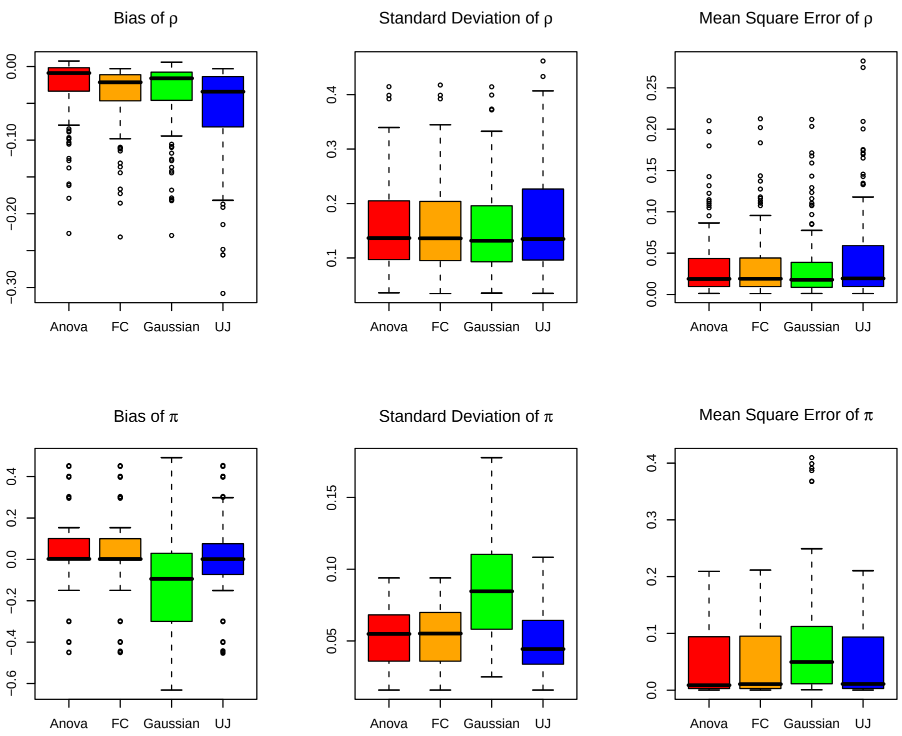

Figure 3.3: The overall performances of the four estimators when $k = 25$

those of the other three estimators. This suggests that ρ_{UJ} is seriously negatively biased in some situations.

The Gaussian estimator ρ_G has the smallest median of SD and MSE while the ANOVA estimator ρ_A has the largest. The 75% quantile of SD and MSE of the new estimator ρ_{UJ} are higher than those of the other three estimators. This suggests that the SD and the MSE of ρ_{UJ} must be larger than the other three estimators in some situations.

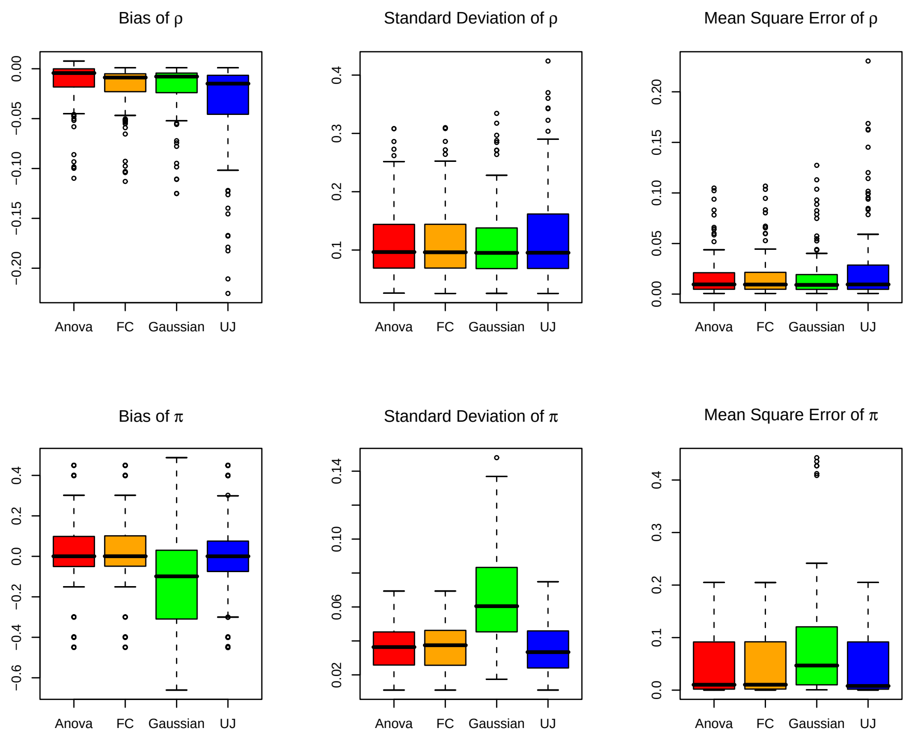

Figure 3.4: The overall performances of the four estimators when $k = 50$

3.2.2 The Effect of the Various Factors

We are also interested in the effects of the various factors on bias, SD and MSE of $\hat{\rho}$. These factors include the sample size (the number of clusters) k, the distribution of the cluster size (n_i), the underlying distribution of S_i and the mean parameter π.

Table (3.2) and Table (3.3) shows the effect of the sample size k, the true value of the mean π, the true value of the correlation ρ and the distribution of cluster size n_i on the bias and MSE of $\hat{\rho}_{UJ}$. From these two tables, we can see that:

- The MSE of $\hat{\rho}$ would increase when the true value of ρ increases and would decrease when the true value of π increases (getting closer to 0.5). So the smallest MSE $\hat{\rho}$ is usually reached at $(\pi, \rho) = (0.5, 0.05)$;

- The MSE $\hat{\rho}$ would decrease when the sample size k increases;
- When all the other factors are fixed, the Brass data would yield higher MSE than that of the Kupper data. As we have shown, "Brass" distribution of n_i has smaller mean than that the "Kupper" distribution of n_i;
- The effects are similar when we compare the bias of the estimators;
- We can get similar conclusion when we look into the estimating results of other estimators.

3.2.3 Comparison Between Different Estimators

Table (3.4) and Table (3.5) are the comparison between the MSE of ρ_{FC} and ρ_{UJ}, for the Kupper data and Brass data respectively. The sample size we use is $k = 25$ and the underlying distribution of S_i is generalized binomial distribution. If the MSE of ρ_{FC} is larger than that of ρ_{UJ}, the font of the cell is bold. Similar results will be obtained when other underlying distribution of S_i and different sample size k are used.

From table (3.4), we can see that for the Kupper data, the MSE of ρ_{UJ} should be smaller than that of ρ_{FC} when π and ρ are both very small (0.05, 0.1 and 0.2). As ρ increases, the MSE of ρ_{UJ} tends to increase more quickly than ρ_{FC} and sometimes would become larger than that of ρ_{FC} when the true value of ρ is large (0.5 and 0.8).

When π increases (getting closer to 0.5), the difference between ρ_{FC} and ρ_{UJ} will become smaller. When $\pi = 0.5$, the MSE of ρ_{UJ} and ρ_{FC} are nearly the same.

From table (3.5), we can see that for the Brass data, the pattern of the change of the MSE is similar with that of the Kupper databut more obvious. For example, when ρ and π are both very small. When $\pi = 0.05$ and $\rho = 0.05$, the MSE of ρ_{FC} is even two times of the MSE of ρ_{UJ} ($\mathbf{0.0421/0.0197}$). However, when $\pi = 0.5$, the MSE of ρ_{UJ} and ρ_{FC} tends to be the same.

Similar results will be got when comparing the bias as well as the standard deviation of these different estimators. We've also found that the properties of the Gaussian estimator ρ_G is close to that of ρ_{UJ} and the properties of the ANOVA estimator ρ_A is close to that of ρ_{FC}.

Figure (3.6), Figure (3.7), Figure (3.8), Figure (3.9) and Figure (3.10) give the relative efficiencies of ρ_A, ρ_G and ρ_{UJ} for different sample size k and different underlying distributions of S_i and n_i. The relative efficiency is defined in (3.1) by using MSE(ρ_{FC}) as the baseline. For each Figure, the left column is based on the Kupper data and the right column is based on the Brass data. For the first row, the underlying distribution is the Beta Binomial distribution and the second row is the "thresholding multivariate normal" distribution we mentioned before. The underlying distribution of S_i for the last row is the generalized binomial distribution. Figure (3.5) is the legend for these figures.

From Figure (3.6), Figure (3.7), Figure (3.8), we can see that:

When $\pi = 0.05$, for the Brass data, the MSE of ρ_{UJ} is larger than those of the other three estimators when the true value of $\rho > 0.5$ but smaller when the true value of $\rho < 0.5$. Here we may call 0.5 the "turning point" for ρ_{UJ}. For the Kupper data, ρ_{UJ} is not obviously better than the other estimators even when the true values of ρ and π are small;

When $\pi = 0.2$, the pattern is similar with the case of $\pi = 0.05$ but not so obvious. And the "turning point" of ρ becomes smaller;

When $\pi = 0.5$, the MSE of ρ_{UJ} is obviously larger than those of the other three estimators when $\rho > 0.2$ and close to the MSE of the other three estimators when $\rho < 0.2$. ρ_{UJ} does not perform better than the other estimators now no matter what distributions of n_i is.

Figure (3.8), Fugure (3.9) and Figure (3.10) shows the effect of k on the MSE when π is small. We can see that for the Brass data, the smaller the k is, the larger the "turning point" of ρ_{UJ} is. Fix $\pi = 0.05$, the "turning point" for the "Brass" data is as following:

In addition, the MSE of the four estimators for the Kupper data are almost the same. Only when $k = 10$, the MSE of ρ_{UJ} is slightly smaller than the other three estimators for some small true values of ρ. We've also found that the effect of the underlying distribution of S_i is so small that we seldom see any differences among the three distributions we use.

3.3 Conclusion

In this chapter, we compared the performances of the four estimators (ρ_{FC}, ρ_A, ρ_G and ρ_{UJ}), for their bias, standard deviation and mean square error. We can make such general conclusions based on the simulation results:

The smaller the true value of ρ is and the closer to 0.5 the true value of π is, the smaller the mean square error of estimator is. The increase of the sample size k will also lead to the decreasing of the mean square error.

The performance of ρ_G is close to that of ρ_{UJ}, but the estimation of π by using the Gaussian method is rather bad, comparing with the U-J method.

For the "Brass" kind data (the mean of the distribution of n_i is small) and small values of π, the MSE of ρ_{UJ} is smaller than those of the other three estimators when the true value of ρ is small but bigger when the true value of ρ is large. The "turning point" is decreasing when the π increases.

For the "Brass" data and small values of π, the turning point of ρ will also increase as k decreases.

We may choose the new estimator ρ_{UJ} in the following situations: the true value of ρ is small, the true value of π is small, the sample size k is small and the mean of the distribution of n_i is small. From the simulation study, we may choose 0.2 as the threshold value of π and 0.5 as the threshold value of ρ.

Table 3.2: The effect of various factors on the bias of the estimator ρ_{UJ} in 1000 simulations from a beta binomial distribution.

			ρ				
π	k	n_i	0.05	0.1	0.2	0.5	0.8
0.05	10	Kupper	-0.0241	-0.0421	-0.0807	-0.2021	-0.1962
		Brass	-0.0501	-0.0672	-0.1143	-0.1760	-0.1597
	25	Kupper	-0.0095	-0.0145	-0.0375	-0.1301	-0.1690
		Brass	-0.0274	-0.0270	-0.0790	-0.1561	-0.1553
	50	Kupper	-0.0012	-0.0100	-0.0199	-0.0659	-0.0963
		Brass	-0.0073	-0.0215	-0.0505	-0.1263	-0.1018
0.10	10	Kupper	-0.0160	-0.0274	-0.0539	-0.1618	-0.1749
		Brass	-0.0519	-0.0731	-0.1031	-0.1590	-0.1528
	25	Kupper	-0.0031	-0.0127	-0.0209	-0.0727	-0.0834
		Brass	-0.0219	-0.0302	-0.0506	-0.1270	-0.1229
	50	Kupper	-0.0029	-0.0048	-0.0065	-0.0298	-0.0368
		Brass	-0.0081	-0.0188	-0.0209	-0.0684	-0.0847
0.20	10	Kupper	-0.0138	-0.0182	-0.0372	-0.0958	-0.1174
		Brass	-0.0541	-0.0639	-0.0785	-0.1429	-0.1323
	25	Kupper	-0.0064	-0.0048	-0.0117	-0.0360	-0.0310
		Brass	-0.0164	-0.0244	-0.0333	-0.0819	-0.0822
	50	Kupper	-0.0027	-0.0036	-0.0081	-0.0125	-0.0088
		Brass	-0.0111	-0.0047	-0.0087	-0.0376	-0.0624
0.50	10	Kupper	-0.0128	-0.0131	-0.0165	-0.0349	-0.0369
		Brass	-0.0581	-0.0414	-0.0618	-0.1139	-0.1086
	25	Kupper	-0.0050	-0.0085	-0.0108	-0.0137	-0.0062
		Brass	-0.0239	-0.0113	-0.0143	-0.0536	-0.0414
	50	Kupper	-0.0023	-0.0057	-0.0059	-0.0038	-0.0027
		Brass	-0.0058	-0.0073	-0.0067	-0.0161	-0.0216

Table 3.3: The effect of various factors on the mean square error of ρ_{UJ} in 1000 simulations from a beta binomial distribution.

			ρ				
π	k	n_i	0.05	0.1	0.2	0.5	0.8
0.05	10	Kupper	0.0041	0.0087	0.0250	0.1180	0.1808
		Brass	0.0188	0.0316	0.0626	0.1732	0.1966
	25	Kupper	0.0020	0.0056	0.0140	0.0835	0.1428
		Brass	0.0124	0.0273	0.0452	0.1330	0.1711
	50	Kupper	0.0013	0.0026	0.0077	0.0444	0.0847
		Brass	0.0085	0.0123	0.0275	0.0993	0.1143
0.10	10	Kupper	0.0038	0.0073	0.0201	0.0944	0.1611
		Brass	0.0260	0.0394	0.0636	0.1461	0.1796
	25	Kupper	0.0018	0.0035	0.0095	0.0432	0.0725
		Brass	0.0121	0.0170	0.0313	0.0972	0.1332
	50	Kupper	0.0009	0.0019	0.0046	0.0182	0.0277
		Brass	0.0062	0.0084	0.0172	0.0527	0.0835
0.20	10	Kupper	0.0037	0.0068	0.0143	0.0563	0.1028
		Brass	0.0273	0.0346	0.0570	0.1254	0.1513
	25	Kupper	0.0014	0.0024	0.0052	0.0181	0.0245
		Brass	0.0099	0.0145	0.0208	0.0611	0.0804
	50	Kupper	0.0007	0.0013	0.0028	0.0066	0.0084
		Brass	0.0058	0.0073	0.0106	0.0269	0.0589
0.50	10	Kupper	0.0034	0.0058	0.0103	0.0216	0.0279
		Brass	0.0269	0.0322	0.0503	0.0917	0.1076
	25	Kupper	0.0013	0.0022	0.0036	0.0075	0.0051
		Brass	0.0100	0.0125	0.0164	0.0390	0.0415
	50	Kupper	0.0007	0.0011	0.0021	0.0035	0.0028
		Brass	0.0050	0.0061	0.0078	0.0147	0.0228

Table 3.4: The MSE of ρ_{FC} and ρ_{UJ} when the cluster size distribution is Kupper

	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$\pi = 0.05$	**0.0147/0.0128**	**0.0267/0.0235**	**0.0546/0.0535**	0.1435/0.1743	0.2126/0.2825
$\pi = 0.1$	**0.0114/0.0106**	**0.0192/0.0181**	0.0361/0.0387	0.0764/0.1078	0.0883/0.1697
$\pi = 0.2$	**0.0064/0.0062**	**0.0104/0.0102**	**0.0186/0.0185**	0.0306/0.0394	0.0189/0.0419
$\pi = 0.5$	0.0025/0.0026	0.0040/0.0041	**0.0075/0.0075**	**0.0112/0.0105**	**0.0076/0.0069**

Table 3.5: The MSE of ρ_{FC} and ρ_{UJ} when the cluster size distribution is Brass

	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$\pi = 0.05$	**0.0421/0.0197**	**0.0656/0.0400**	**0.0956/0.0735**	0.1834/0.2095	0.2019/0.2746
$\pi = 0.1$	**0.0270/0.0176**	**0.0413/0.0305**	**0.0635/0.0570**	0.1133/0.1650	0.0901/0.2003
$\pi = 0.2$	**0.0182/0.0155**	**0.0220/0.0214**	0.0364/0.0382	0.0486/0.0888	0.0322/0.1179
$\pi = 0.5$	0.0114/0.0116	0.0140/0.0150	0.0179/0.0192	0.0222/0.0439	0.0140/0.0681

Table 3.6: The "turning point" of ρ when $\pi = 0.05$

	k=10	k=25	k=50
turning point of ρ	0.5	0.4	0.3

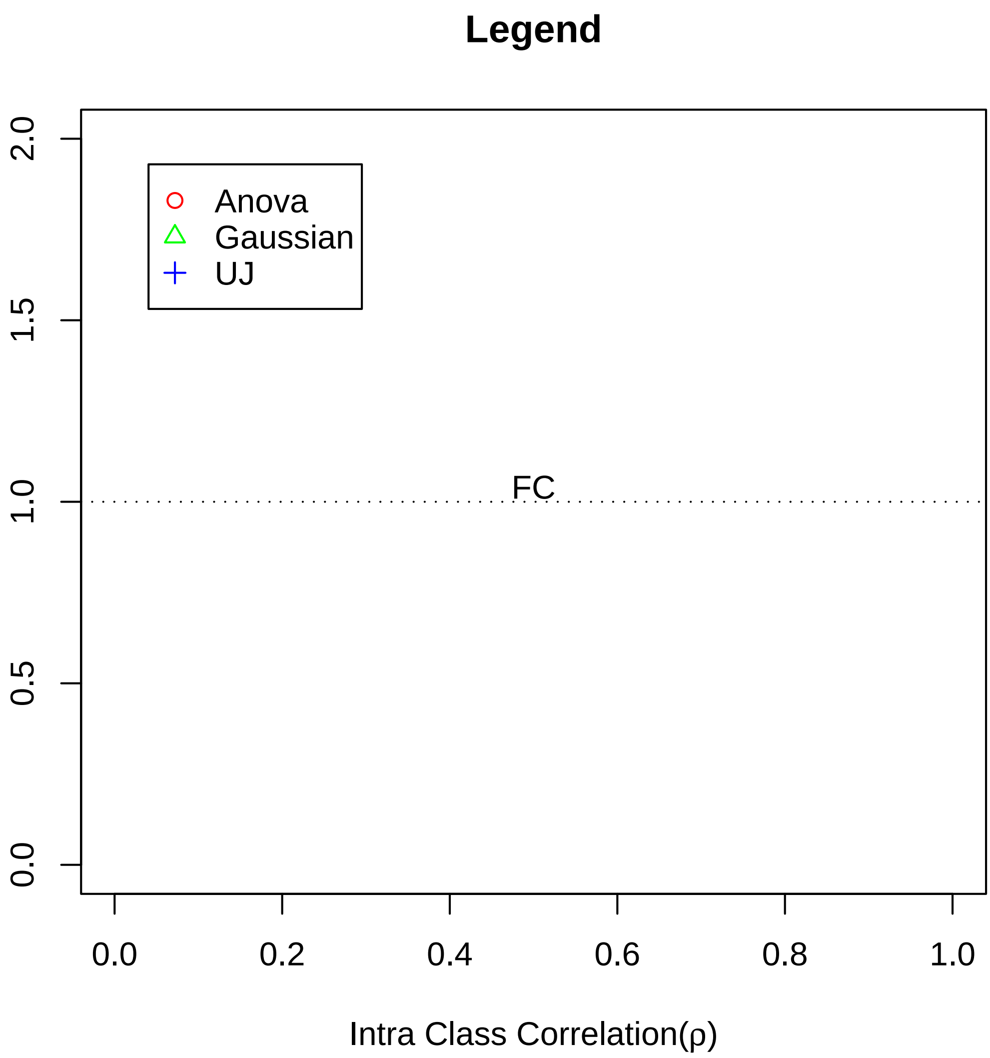

Figure 3.5: The Legend for Figure (3.8), (3.7), (3.6), (3.9) and (3.10)

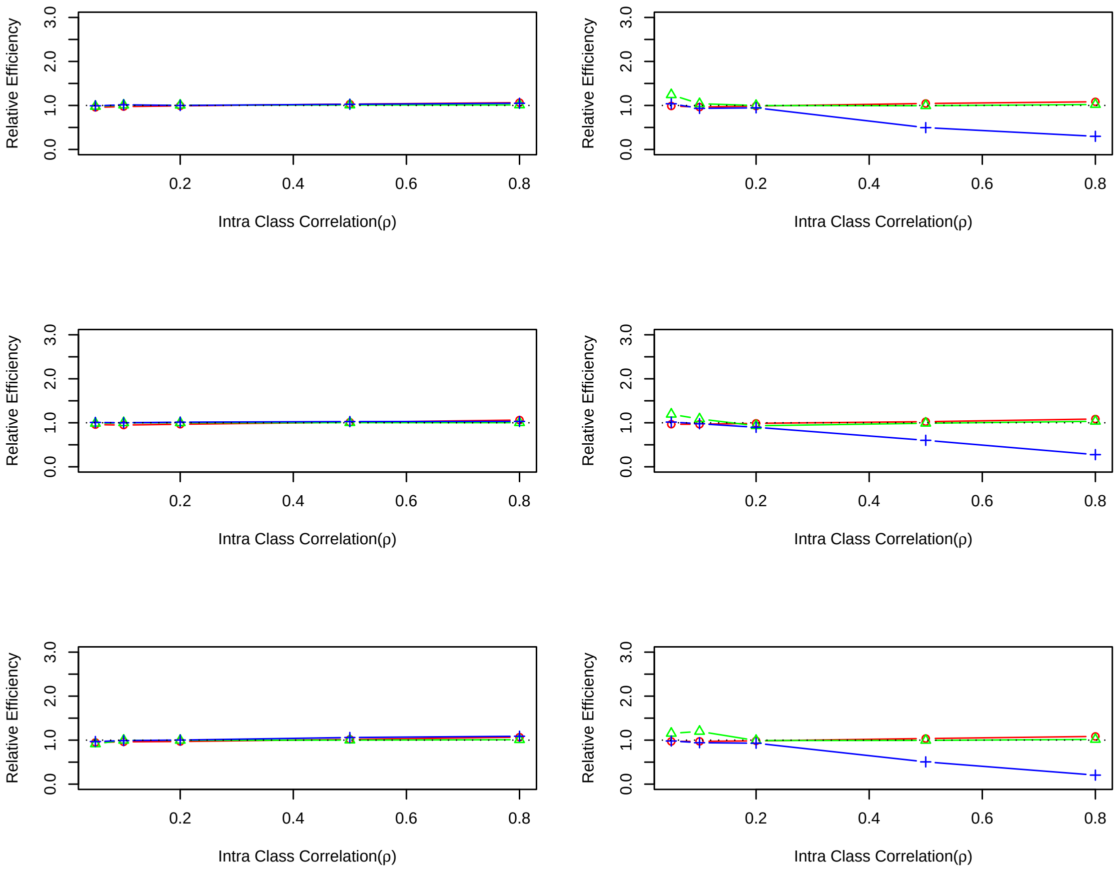

Figure 3.6: The Relative Efficiencies when $k = 25$ and $\pi = 0.5$

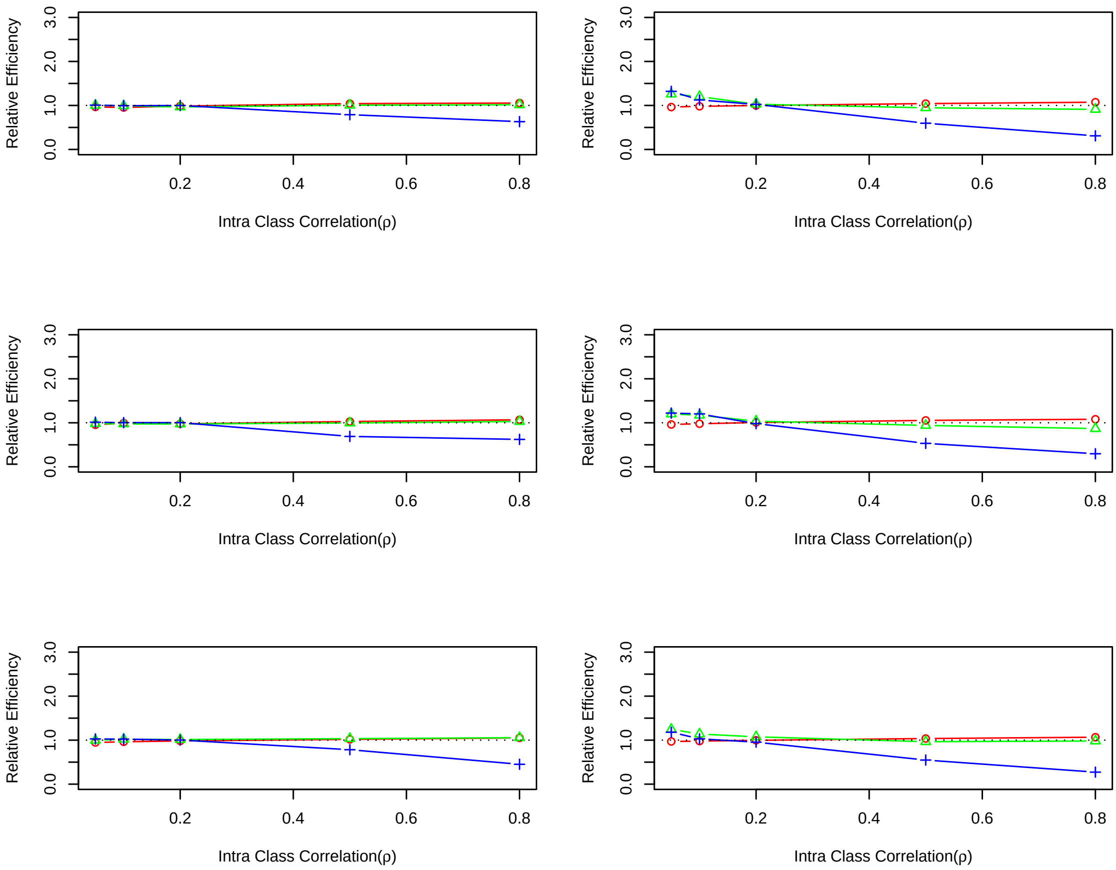

Figure 3.7: The Relative Efficiencies when $k = 25$ and $\pi = 0.2$

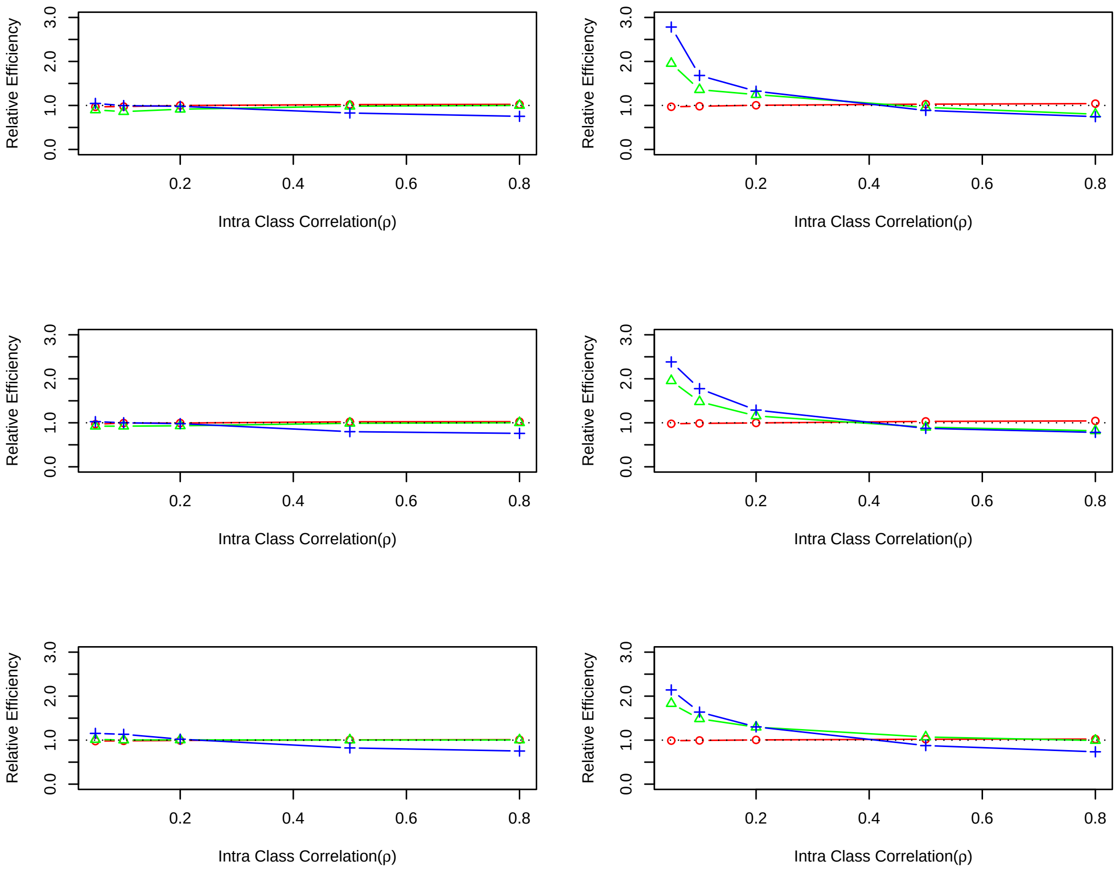

Figure 3.8: The Relative Efficiencies when $k = 25$ and $\pi = 0.05$

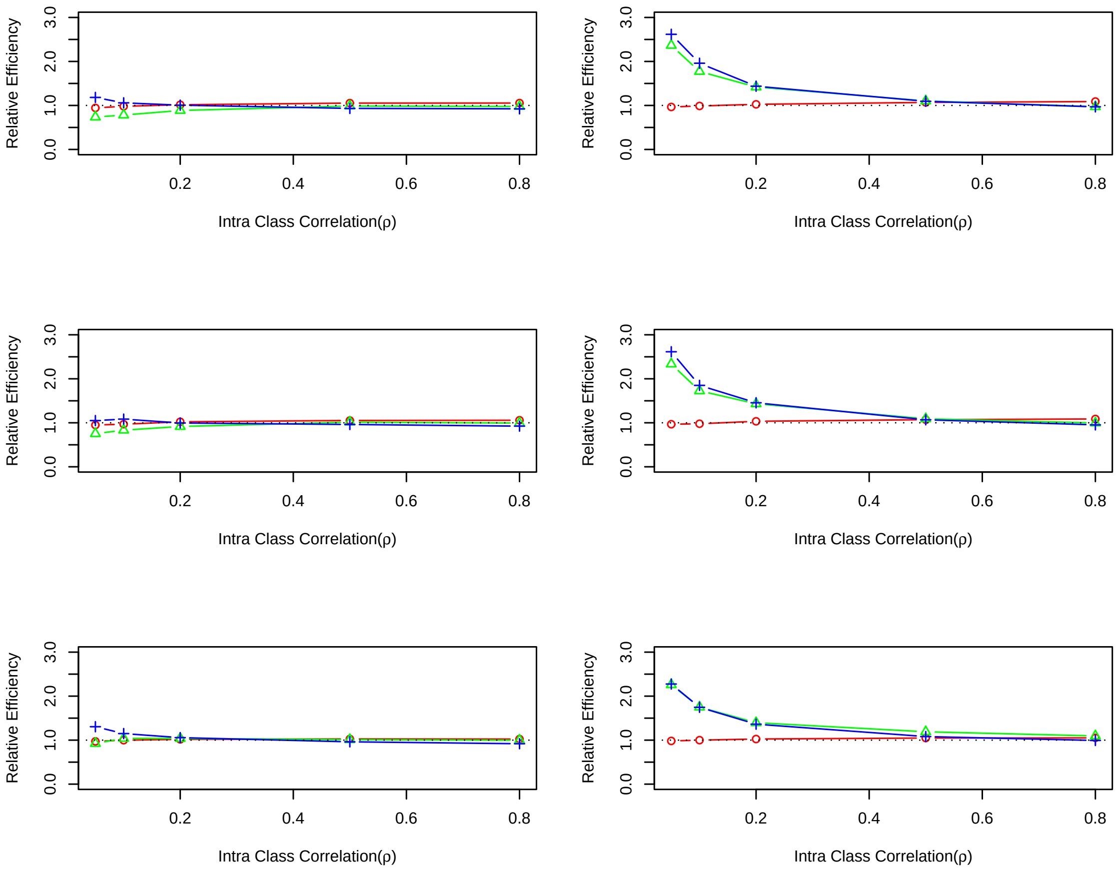

Figure 3.9: The Relative Efficiencies when $k = 10$ and $\pi = 0.05$

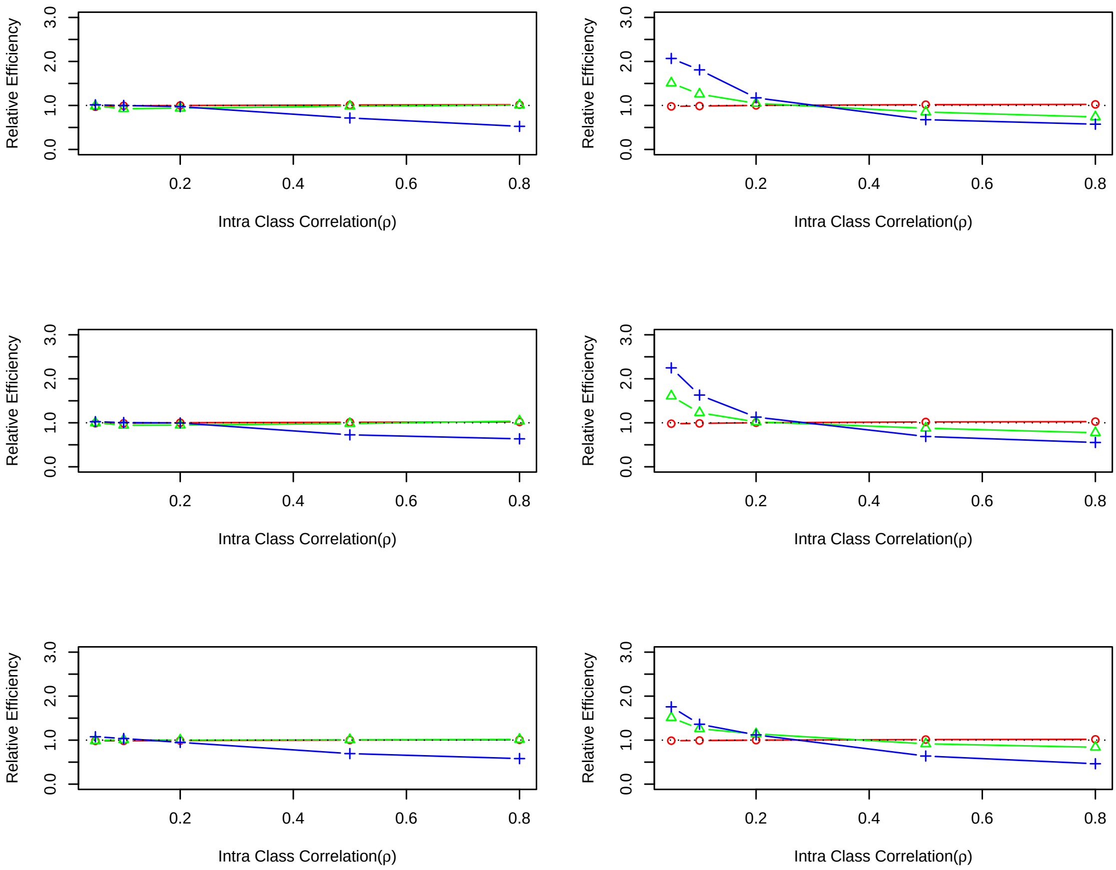

Figure 3.10: The Relative Efficiencies when $k = 50$ and $\pi = 0.05$

Chapter 4

Real Examples

4.1 The Teratological Data Used in Paul 1982

The first data set we use is from the Shell Toxicology Laboratory, which was first used by Paul (1982). It is a typical teratology study data, which contains a control group and three treatment groups. For different groups, they are supposed to have different means and we are interested in the intra-group correlation within each group. Table(4.1) shows the data structure:

4.2 The COPD Data Used in Liang 1992

The second data set we use is the COPD data from Liang et al.(1992). The familial aggregation of the Chronic Obstructive Pulmonary Disease (COPD) is used as a measure of how genetic and environmental factors may contribute to disease etiology. It involves 203 siblings from 100 families. The binary response here indicate whether a

Table 4.1: Shell Toxicology Laboratory, Teratology Data

Group		
Control	s_i	1 1 4 0 0 0 0 0 1 0 2 0 5 2 1 2 0 0 1 0 0 0 0 3 2 4 0
	n_i	12 7 6 6 7 8 10 7 8 6 11 7 8 9 2 7 9 7 11 10 4 8 10 12 8 7 8
Low dose	s_i	0 1 1 0 2 0 1 0 1 0 0 3 0 0 1 5 0 0 3
	n_i	5 11 7 9 12 8 6 7 6 4 6 9 6 7 5 9 1 6 9
Medium dose	s_i	2 3 2 1 2 3 0 4 0 0 4 0 0 6 6 5 4 1 0 3 6
	n_i	4 4 9 8 9 7 8 9 6 4 6 7 3 13 6 8 11 7 6 10 6
High dose	s_i	1 0 1 0 1 0 1 1 2 0 4 1 1 4 2 3 1
	n_i	9 10 7 5 4 6 3 8 5 4 4 5 3 8 6 8 6

sibling of a COPD patient has impaired pulmonary function. Table (4.2) shows the data structure.

Table 4.2: COPD familial disease aggregation data

Siblings	1	1	2	2	2	3	3	3	3	4	4	4	6	6	6	6	6
COPD Patients	0	1	0	1	2	0	1	2	3	0	1	2	0	2	3	4	6
Families	36	12	15	7	1	5	7	3	2	3	3	1	1	1	1	1	1

Take the last column for example: there are one such family, of all the 6 siblings in the family, 6 siblings are the COPD patients.

4.3 Results

We use five data sets in our ”Real Example” section. Four data sets from the Teratology data used by Paul (1982) and one data set from the COPD data used in Liang (1992). Table (4.3) shows the estimating results of these five data sets.

Table 4.3: Estimating Results for the Real Data Sets

	Control		Low Dose		Medium Dose		High Dose		COPD	
	π	ρ	π	ρ	π	ρ	π	ρ	π	ρ
FC	0.1409	0.2091	0.1280	0.0916	0.3458	0.2636	0.2385	0.1371	0.2823	0.1800
Anova	0.1410	0.2189	0.1274	0.1030	0.3458	0.2780	0.2392	0.1531	0.2821	0.1855
Gaussian	0.0471	0.2262	0.1214	0.0972	0.3159	0.2723	0.2038	0.1389	0.2604	0.2074
UJ	0.1409	0.2123	0.1286	0.1138	0.3459	0.3056	0.2379	0.1238	0.2946	0.2209

From Table (4.3), we can see that the estimating results of the four estimators of ρ are almost the same. But for the Gaussian estimator, the estimating result for π is much different from the other three estimators, which is consistent with the finding of the simulation study in Chapter 4.

Based on the findings of the simulation study, we know that when the true values of π are small (usually using 0.2 as the threshold value), the UJ method has smaller MSE than those of the other estimators. In our case, we can rely on the UJ method for the control and low dose group (the true value of π are believed to be smaller than 0.2). But for the other groups of data, we can not guarantee that the UJ method is better. We have to compare the asymptotic variance of these estimators, by using the methods we discussed in Chapter 2.

By plugging in $(\hat{\pi}, \hat{\rho})$ we obtained into the formula (2.26) and (2.21), (2.29), (2.28), we can get the estimated values of the asymptotic variances of ρ_G, ρ_{UJ}, ρ_{FC} and ρ_A. Table (4.4) is the results for our data sets.

Note that many of the estimated values of the asymptotic variances in Table (4.4) are negative. As mentioned in Chapter 2, when the sample size k is large, it would be fine to simply plug in $(\hat{\pi}, \hat{\rho})$ we obtained into the theoretical formulas. However, for our

Table 4.4: The Estimated value of the Asymptotic Variance of $\hat{\rho}$ (By plugging the estimates of (π, ρ)) into formulas: (2.29), (2.28), (2.26) and (2.21)

	Control	Low Dose	Medium Dose	High Dose	COPD
FC	-0.0066	-0.0090	0.0007	-0.0040	-0.03339
Anova	-0.0068	-0.0099	0.0008	-0.0053	-0.0336
Gaussian	-0.0870	-0.0137	-0.0010	-0.0057	-0.0457
UJ	-0.0075	-0.0084	0.0050	-0.0050	-0.0109

data sets, the k is not large enough to avoid encountering the negative value of the estimated asymptotic variances. So robust methods mentioned in Chapter 2 should be used. Table (4.5) is the estimated values of the asymptotic variances by using the robust method.

Table 4.5: The Estimated value of the Asymptotic Variance of $\hat{\rho}$ (by using the Robust Method)

	Control	Low Dose	Medium Dose	High Dose	COPD
FC	0.0056	0.0066	0.0169	0.0174	0.0186
Anova	0.0058	0.0070	0.0174	0.0183	0.0186
Gaussian	0.0776	0.0074	0.0145	0.0197	0.0196
UJ	0.0048	0.0066	0.0123	0.0131	0.01589

From Table (4.5), we can see that the estimated asymptotic variance of ρ_{UJ} are smaller than the other three estimators, no matter which real data set we concerned. Thus we can say that we may choose ρ_{UJ} to estimate the ICC in the above two data sets and the estimating results are reliable.

Chapter 5

Future Work

We are now supposing that the mean parameter for each cluster is the same, that is $\pi_i = \pi$, for any $i = 1, 2, \ldots, k$. Actually, the π_i may be different for different clusters. When ρ is close to 1 or the variance of π_i is small, the common mean parameter π we use can be considered to be the expected value of the mean parameter π_i; otherwise, this approximation maybe inappropriate. In our future work, we will investigate the properties of the estimating equations when π_i are different.

Another work we shall do is to generalize the estimating functions for intra-class correlation parameter ρ. After soem algebra, the Gaussian estimating function (2.5) can be written as: $g_G(\rho) = \sum_i \varepsilon_i^T M_i \varepsilon_i$ where $M_i = I_i - \frac{1+(n_i-1)\rho^2}{(1+(n_i-1)\rho)^2} J_i$ (I_i is the unit matrix and J_i is the matrix constituted of 1s). And the UJ estimating function (2.6) can also be written as: $g_J(\rho) = \sum_i \varepsilon_i^T M_i \varepsilon_i$ where $M_i = \frac{C_i}{n_i(n_i-1)} \Big\{ [1 + (n_i - 1)\rho] I_i - J_i \Big\}$ (C_i as defined in (2.6)). This motivates us to find a general form of the estimating functions $g(\rho) = \sum_i \varepsilon_i^T M_i \varepsilon_i$, where M_i is the linear combination of I_i and J_i. We will try to find the best M_i to maximize the efficiency of the estimation of ρ.

Furthermore, we can even extend the result to the general longitudinal data in which the response may be continuous and the correlation matrix may not be compound symmetry.

Bibliography

Agresti, A (1990). *Categorical Data Analysis.* New York: John Wiley and Sons.

Carey, V., Zeger, S. L, and Diggle, P. (1993). Modeling multivariate binary data with alternative logistic regressions. *Biometrica*, **80**, 517-526.

Crowder, M. J. (1979) Inference about the intraclass correlation coefficient in the beta-binomial ANOVA for proportions. *J. R. Statist. Soc. B*, **41**, 230-234

Crowder, M. (1985). Gaussian Estimation for Correlated Binomial Data. *Journal of Royal Statistical Society* B, **47,** 229–237.

Crowder, M. (1987). On linear and quadratic estimating equations. *Biometrica*, **74**, 591-597.

Donner, A. (1986) A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *Int. Statist. Rev.*, **54**, 67-82.

Elston, R.C. (1977). Response to query, consultants corner. *Biometrics* **33**, 232–233.

Feng, Z. and Grizzle, J. E. (1992) Correlated binomial variates: Properties of estimator of Intra-class correlation and its effect on sample size calculation. *Statistics in Medicine*, **11**, 1607-1614

Fleiss, J. L. and Cuzick, J. (1979) The relaibility of dichotomous judgements: unequal number of judges per subject. *Appl. Psychol. Bull.*, **86**, 974-977.

Landis, J. R. and Koch, G. G. (1977a) A one-way components of variance model for categorical data. *Biometrics*, **33**, 671-679.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Liang, K.Y. and Hanfelt, J. (1994). On the use of the Quasi-likelihood method in teratological experiments. *Biometrics* **50**, 872–880.

Lipsitz, S.R. and Fitzmaurice, G.M. (1996). Estimating equations for measures of association between repeated binary responses. *Biometrics* **52,** 903–12.

Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78,** 153–60.

Kupper, L.L. and Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* **35**, 281–293.

Madsen, R. W. (1993). Generalized binomial distributions. *Communications in Statistics, Part **A**–Theory and Methods*, **22**, 3065-3086.

Mak, Tak K. (1988). Analysing Intraclass Correlation for Dichotomous Varibales . *Applied Statistics*, **37**, 344-352.

Paul, S.R. (1982). Analysis of proportions of affected foetuses in teratological experiments. *Biometrics* **38,** 361–370.

Paul, S. R. and Islam, A. S. (1998). Joint estimation of the mean and dispersion parameters in the analysis of proportions: a comparison of efficiency and bias. *The Canadian Journal of Statistics*, **26**, 83C94.

Paul, S.R. (2001). Quadratic estimating equations for the estimation of regression and dispersion parameters in the analysis of proportions. *Sankhya*, **63**, 43–55.

Paul, S.R., Saha, K.K. and Balasooriya, U. (2003). An empirical investigation of different operation characteristics of several estimators of the intraclass correlation in the analysis o binary data. *J. Statist. Comp. Simul.* **73,** 507–523.

Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association*, **81**, 321–327.

Ridout, M.S., Demétrio, C.G.B. and Firth, D. (1999). Estimating intraclass correlation for binary data. Biometrics, **55**, 137–148.

Wang, Y.-G. and Carey, V.J. (2003). Working correlation structure misspecification, estimation and covariate design: implications for GEE performance. *Biometrika* **90,** 29–41.

Wang, Y.-G. and Carey, V.J. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *J. Amer. Statist. Assoc.* **99,** 845-853.

Zeger, S.L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.

Zhu, Min (2004). Overdispersion, Bias and Efifciency in Teratology Data Analysis. *A thesis submitted for the degree of Master of Science, Department of Statistics and Applied Probability, National University of Singapore*

Zou, G. and Donner, A. (2004). Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. *Biometrics*, **60**, 807–811.